

Brown's budget briefing

A more cohesive biomedical research agency and simpler arrangements for measuring university performance feature in Britain's 2006 budget.

In most years, in most countries, the only element in a finance minister's budget statement that grabs researchers' immediate attention is the measly percentage increase that their funding agency has managed to prise out of government for the coming year.

But when Gordon Brown, the British chancellor of the exchequer, stood up to deliver his budget statement on 23 March, he had some interesting things to say to scientists. The statement proposed two reform initiatives that most researchers will welcome, if they are implemented with care.

Thanks in part to Brown's conviction that research is key to economic growth, the Labour government has upped annual science spending by about 70% since it came to power in 1997, to more than £2.5 billion (US\$4.4 billion). There were no spending increases this time. But Brown, who is expected to succeed Tony Blair as prime minister at some point in the next two years, has clearly been giving some thought to how British research could be strengthened.

The most significant planned change is a shake-up of biomedical research, to bring science and clinical trials under one roof. At present, they are divided between the Medical Research Council (MRC), which supports biomedical science, and the National Health Service (NHS), which runs clinical trials and other healthcare research through its network of local health trusts.

Two into one

The NHS research and development budget is £750 million, around £200 million more than that of the MRC, but NHS research has a far lower profile, both scientifically and publicly. One reason is that the money is spent through the local trusts, which also provide day-to-day healthcare. In this situation, it is perhaps inevitable that some resources theoretically allocated to long-term research end up diverted to immediate healthcare needs.

Brown aims to end this split. A merger of the NHS's research with that of the MRC will create a new, as yet unnamed agency, which the Treasury says will receive "at least" £1 billion in funding each year. The agency will support all biomedical research, from basic studies to clinical trials, in the manner of the US National Institutes of Health. Senior researchers have welcomed the plan. Even if the annual budget is slightly less than the total budget of the two current operations, it will still represent a significant increase in biomedical research funding, as all of the money will now actually be spent on research.

The merger will be complex to implement, however. The budget statement says that the new arrangement will be jointly administered by the Department of Health, which runs the NHS, and the Department of Trade and Industry (DTI), which currently oversees the MRC and Britain's other research councils. But the research councils enjoy arms-length relationships with the DTI that prevent political meddling and protect their scientific reputations. It is vital that the

new body has a similar arrangement to protect it from political interference by ministers at either of the departments that will be responsible. The government has yet to clarify how this will be achieved.

Researchers will also be glad to see the back of the Research Assessment Exercise, a mechanism to measure university departments' performance, whose expiry after a final outing in 2008 was confirmed in Brown's budget statement. The exercise, which has taken place every seven years or so since 1985, has helped the government to determine the levels of fixed funding, on top of research grants, for Britain's universities. But it is vastly time-consuming for overseers and overseen alike, and its usefulness as a management tool has been gradually waning.

Measure for measure

Brown is proposing to replace the Research Assessment Exercise with a system that rewards departments on the basis of performance metrics. One metric highlighted in the budget statement, external research income, seems a reasonable basis for departmental funding, as this income correlates very well with the peer review that was done in the assessment exercise.

But other types of research metrics — however attractive they might look to those who make funding choices — should be handled with great care. Citation statistics, for example, are a notoriously unreliable and inconstant guide to research quality. The top research agencies in the United States have successfully resisted periodic drives by bureaucrats to use them to measure the worth of the science that they should be supporting.

One way forward for the UK government would be to let reliable metrics, such as total external research income, replace much of the Research Assessment Exercise process, while retaining a slimmed-down version of the subject panels that currently oversee each discipline. All disciplines could then base their evaluations on research income, but the subject panels could tweak the process to reflect the needs of their particular fields of study. For example, they might choose to reward academics working on valuable long-term projects, such as the collection of environmental or astronomical data sets, the importance of which is not reflected in the metrics.

The government will now consult interested parties on both the biomedical and research-assessment proposals, each of which holds considerable potential. If the scientific community has a voice in how the proposals are implemented, that potential will be realized. There might not be new money on the table in this particular budget, but it contains ample opportunity to build a better environment for science in Britain.

"It is vital that the new biomedical-research body has an arrangement to protect it from political interference by ministers."



Astrobiology at ten

A young discipline holds promise yet.

In the 1960s, the evolutionary theorist George Gaylord Simpson rubbished the then-nascent science of exobiology, which concerned itself with life on places other than Earth, as a science without a subject.

Now the science descended from exobiology, astrobiology, is heading towards its tenth birthday. But as hundreds of its practitioners assemble this week in Washington for their annual AbSci-Con meeting, the same criticism is being heard once more. And simpsonian scepticism that astrobiology has any useful data or insights to offer makes it easy to cut budgets for the discipline, as NASA has done (see *Nature* **439**, 768–769; 2006). As we report on page 586 of this issue, these cuts have now been partially rescinded, but the discipline's future remains cloudy.

Setting aside for now the difficulties of allocating a constrained NASA science budget, the fundamental scepticism voiced by some of astrobiology's critics is misplaced, for at least two reasons. The first is the timescale of space science. Ten years is no time at all in terms of mission planning — and as a result there has not yet been a single astrobiological space mission. The first two that might make it off the launch pad will get under way in the next couple of years: the Kepler mission, which will look for planets the size of Earth circling other stars in orbits that might allow liquid water at the surface, and the Phoenix lander, which will look for organic molecules in ice on Mars. Both are important missions, Kepler profoundly so.

The second reason is a misunderstanding — sometimes wilful, sometimes not — concerning the nature of astrobiology itself. Although the field was cooked up, in part, out of political necessity, as a means of bundling together research programmes on exobiology, other life sciences and planetary science, it has at its core a powerful unifying idea. Whereas exobiology specifically took life elsewhere in the Universe as the object of its study, astrobiology looks at life in the context of the Universe — in the context, that is, of astronomy and planetary science. Thus astrobiology legitimately broadens the terms of exobiology to include the study of life on Earth, which in this context is just another planet — albeit one to which we enjoy privileged access.

The inclusion of studies of life on Earth in astrobiology has provided opportunities to re-brand existing work. It appears that many microbiologists with an interest in extremophile microbes have suddenly become astrobiologists, because astrobiology is — or was — where the money is. But it has also provided the field with a coherence that exobiology always lacked.

Life arises in an astronomical context, and Earth itself is part of that context. It is an intersection of the local and the cosmic, of the deepest time and the newest intelligence. Astrobiology has given this perspective an institutional home. Some second-rate research may have been funded on occasion, thanks to the astrobiology moniker's modishness. But the science indeed has a subject. It is a powerfully evocative one, which resonates not just with scientists, but with a wider public as well. ■

"Ten years is no time in terms of mission planning — and there has not yet been a single astrobiology space mission."

Britannica attacks

... and we respond.

Last December, *Nature* published a News story about the accuracy of two online references sources. We compared the website of an established publication, Encyclopaedia Britannica, with that of Wikipedia, a new kind of online encyclopaedia that anyone can edit and update, regardless of expertise.

The result (see *Nature* **438**, 900–901; 2005) surprised us, and many others. Forty-two expert reviewers carried out the comparison. After we had tallied their results, we saw that they had picked up errors (the great majority of them minor) at a rate of about three per online Britannica item and about four per Wikipedia item.

Last week, Encyclopaedia Britannica issued a statement (http://corporate.britannica.com/britannica_nature_response.pdf), and this week published a half-page advertisement in the London *Times* criticizing our study and demanding that we retract our story.

Britannica complains that we did not check the errors that our reviewers identified, and that some of them are not errors at all. We disagree with their claims in some of the cases (others are too specialized for an immediate response), but there is a more important point to make. Our reviewers may have made some mistakes — we have been open about our methodology and never claimed otherwise — but the entries they reviewed were blinded: they did not know which entry came from Wikipedia and which from Britannica. We see no reason to believe that any misidentifications of errors would adversely affect one publication more than the other. And of

the 123 purported errors in question, Britannica takes issue with fewer than half.

Another Britannica criticism concerns the fact that we provided material from other Britannica publications, such as the Britannica Book of the Year. This was deliberate: the aim of our story, as we made clear, was to compare the online material available from Britannica and Wikipedia. When users search Britannica online, they get results from several Britannica publications. They have no reason to think that any one is less reliable than the others. In the case of some year-book entries, Britannica itself asks readers to reference the articles as coming from "Encyclopaedia Britannica Online" — exactly the source we set out to compare.

Other objections are simply incorrect. The company has, for example, claimed that in one case we sent a reviewer material that did not come from any Britannica publication. When the company made this point to us in private we asked for details, but it provided none. Now Britannica has identified the review in question as being on ethanol. We have checked the original e-mail that we sent to the reviewer who looked at the Britannica article on ethanol, and it is clear to us that all the reviewer's comments refer to specific paragraphs from Britannica.

Our responses to the points raised by Britannica in its original online posting and in its subsequent advertisement can be found at <http://www.nature.com/nature/britannica/index.html>. Our comparison was unbiased, and we reject Britannica's allegation that we have acted in a dishonest manner. We stand by the story. ■

"Entries were blinded — reviewers did not know which entry came from Wikipedia and which from Britannica."

RESEARCH HIGHLIGHTS

Mapping the spread

Emerg. Infect. Dis. [online] www.cdc.gov/ncidod/EID/vol12no02/05-0640.htm (2006)

As the H5N1 avian flu virus advances across Europe and Africa, a study of detailed outbreak data from Thailand shows how researchers can use geospatial analysis to understand what drives the disease's spread.

Marius Gilbert of the Université Libre de Bruxelles in Belgium and his colleagues looked at how the spatial distribution of H5N1 outbreaks in birds related to land use. They found surprisingly little association between disease distribution and the distribution of backyard chickens. Instead, free-ranging ducks feeding in paddy fields emerged as the greatest risk factor, suggesting that H5N1 control strategies in Thailand should focus on controlling the movement of ducks among rice paddies.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

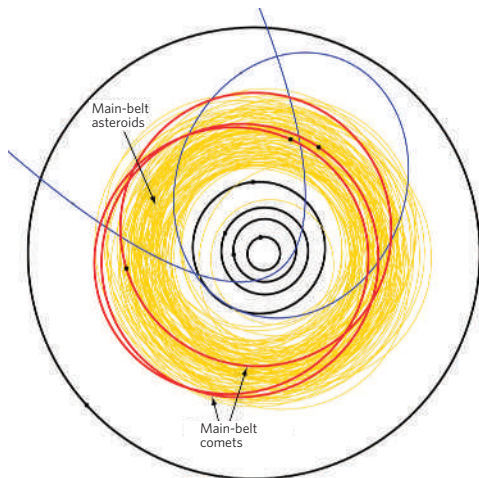
ASTRONOMY

Catch a comet by its tail

Science doi:10.1126/science.1125150 (2006)

Astronomers have new evidence for a third population of comets, closer to home than the two known reservoirs at the fringes of the Solar System.

Henry Hsieh and David Jewitt, both from the University of Hawaii in Honolulu, studied around 300 objects in the main asteroid belt. The diagram below shows the paths of some of the asteroids (yellow), which orbit between Mars and Jupiter (black). The researchers spotted three objects that are shedding dust, as do comets frazzled by the sun (red). They argue that the objects are comets that formed in the region and were later activated by collisions. The team estimates that there may



be up to 150 such active comets in the main asteroid belt. The blue lines show the paths of two conventional comets, Halley and Tempel 1, which originated in the more distant populations.

GENETICS

Fishing for key sequences

Science doi:10.1126/science.1124070 (2006)

Geneticists commonly track down important gene-control regions of DNA by searching for sequences that are similar in widely differing species. But this technique may overlook many key sequences, according to Shannon Fisher and Andrew McCallion at the Johns Hopkins School of Medicine in Baltimore, Maryland, and their colleagues.

The team shows that human DNA sequences thought to control a central developmental gene called *RET* also correctly direct the counterpart gene in the zebrafish (*Danio rerio*), even though the sequences from the two species bear no obvious similarities. The sequences may retain the ability to bind regulatory proteins despite their divergence, the researchers propose.

ATOMIC PHYSICS

Atoms fall into the trap

Phys. Rev. Lett. (in the press); preprint at <http://arxiv.org/abs/physics/0603157> (2006)

A promising route to a quantum computer involves storing quantum bits of information on atoms held in microfabricated traps. Trapping charged ions on chips using

electromagnetic fields is now a fine art. But neutral atoms are preferable, as they interact less with their environment — a process that can let quantum information leak away.

Hidetoshi Katori of the Japan Science and Technology Agency in Tokyo and his colleagues have now held nearly 100 neutral atoms of strontium in a microscopic electrical trap. Unlike previous neutral-atom traps, this one doesn't rely on magnetic interactions — the strontium atoms have no magnetic moment — and so is not susceptible to stray magnetic fields. The atoms stay put for up to 80 microseconds.

BIOCHEMISTRY

Four's a crowd

Nature Methods 3, 267–273 (2006)

A partnership based on one of the strongest non-covalent bonds in nature has been improved. Biochemists and nanotechnologists working with the protein streptavidin — which binds strongly to biotin — may no longer need to worry about their mixtures forming a sticky mess.

Streptavidin is used to track or anchor biotin-labelled molecules, but because it is a tetramer, it can bind four biotins at once. This leads to clumping. To counter the problem, Alice Ting of the Massachusetts Institute of Technology in Cambridge and her group created a streptavidin tetramer with only one biotin binding site. They used the manufactured tetramer to track biotinylated neuroligin-1, a neuronal adhesion protein that clumps when the normal tetramer is used.

NEUROBIOLOGY

Garlic sensation*Cell* **124**, 1269–1282 (2006)

The pungent ingredients of mustard oil and garlic are known to activate the TRPA1 receptor, a member of the TRP family of ion channels that detect painful stimuli. David Julius at the University of California at San Francisco and his colleagues build on their earlier characterization of this receptor in isolated sensory neurons, with studies in mice.

The researchers report that mice lacking TRPA1 were completely insensitive to the pungent compounds in mustard oil and garlic, suggesting that TRPA1 is the compounds' sole target. The TRPA1-deficient mice were also less responsive to environmental irritants such as acrolein, an ingredient of tear gas and vehicle exhaust. However, the mice were normal in sensing extreme cold and loud sounds, contrary to suggestions that TRPA1 has a crucial role in these processes.

MEDICINE

Healthy diet, happy heart*N. Engl. J. Med.* **354**, 1264–1272 (2006)

A lifetime of low cholesterol could slash the risk of heart disease, say Helen Hobbs of the University of Texas Southwestern Medical Center, Dallas, and her colleagues.

The team examined a group of people who carry rare versions of the PCSK9 gene that help to rid the body of low-density lipoprotein (LDL) cholesterol. The group was up to 88% less likely to develop coronary heart disease over the 15 years of the study than those carrying more common gene variants.

The results suggest that modest, lifetime reductions in LDL cholesterol — which may be achievable by exercise, diet or drugs — could reduce the rate of heart disease far more effectively than combating cholesterol at middle age.

CELL BIOLOGY

MicroRNAs under suspicion*Cell* **124**, 1169–1181 (2006)

MicroRNAs — short RNAs that control gene function — were discovered 13 years ago, but there are still few clues to their precise roles.

Reuven Agami from The Netherlands Cancer Institute in Amsterdam and his colleagues have now completed a genetic screen of known human miRNAs. They looked for miRNAs that may induce cancer, identifying two that interfere with the tumour-suppressing p53 system.

The researchers then found that these two

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

miRNAs were expressed in all tested samples of testicular germ-cell tumours that had non-mutated p53 genes, confirming their clinical relevance.

GEOLOGY

Continents rock*Geology* **34**, 245–248 (2006)

The young Earth had a more widespread continental crust than was once thought, a new study suggests.

Radiometric dating of a zircon crystal, found in the ancient metamorphic rocks of northwestern Canada, indicates that continental crust was present there 4.2 billion years ago — just a few hundred million years after the Earth had formed. It is the only indication outside Australia, where similarly ancient zircons have been found, that granite-rich continental crust existed so early in the planet's history.

Such crust may have covered a significant portion of the early Earth, say Tsuyoshi Iizuka of the Tokyo Institute of Technology, Japan, and his team.

CHEMICAL BIOLOGY

A taste for DNA*Angew. Chem. Int. Edn* **45**, 2238–2242 (2006)

A single strand of DNA is all it takes, in principle, to activate the 'detection machine' devised by Itamar Willner and his colleagues at the Hebrew University of Jerusalem, Israel. Once triggered, the device chomps on strands of 'fuel' DNA, releasing fluorescent dye as it goes.

The machine is a DNA-cleaving enzyme; it is switched on when the target DNA strand pairs up with a complementary sequence looped around the enzyme's jaws. The enzyme then munches through the fuel DNA, releasing fluorescent marker as a waste product, and reconstructing itself before each fresh cycle of cleavage.

The researchers demonstrate their technique on a sequence associated with the Tay–Sachs genetic disorder, and say that it might provide an alternative to PCR-based sequence detection.

JOURNAL CLUB

Gordon McFiggans

University of Manchester, UK

An atmospheric scientist ponders a cloudy problem.

Reluctantly, I was forced to open an old can of worms when writing a review article on cloud formation.

The problem concerns how water droplets grow. Numerical models of cloud formation assume that any water-vapour molecules colliding with a liquid droplet will be accommodated on to the droplet's surface.

However, several groups have reported measurements that are inconsistent with this argument. These measurements suggest that something inhibits the uptake of water — possibly some component of the aerosol particle that originally 'seeded' the water droplet, which coats the growing droplet's surface.

The implications of this debate are more than academic. Predicting how clouds influence climate requires a good model of cloud brightness and persistence — properties that depend on droplet number and thus, ultimately, uptake probability.

My attention was drawn to work from the Queensland University of Technology in Brisbane, Australia, where researchers had used a home-built instrument to investigate aerosol particles. Their instrument sized the particles, then heated them to drive off volatile material before increasing the humidity. Finally, the instrument measured the size of the resulting water droplet (G. Johnson *et al.* *J. Geophys. Res.* **110**, D20203; 2005).

A subset of the particles took up more water after heating, suggesting the lost volatile component had inhibited water uptake. This is the most direct evidence I have seen for a mechanism for suppressed water uptake in the real atmosphere.

In our review article, we dealt with the debate by laying out the arguments without drawing a clear conclusion. But experiments such as those done by Johnson *et al.* may yet provide a definitive answer.

NEWS

Stem cells from testes: could it work?

On Friday 24 March, researchers at the Georg August University of Göttingen in Germany announced they had found a source of reprogrammable cells in the testes of adult mice. By Sunday morning, more than ten companies had e-mailed senior team member Gerd Hasenfuss about collaborating on future work.

Hasenfuss, a cardiologist, admits that much work must be done before the discovery could lead to any applications in humans. But the finding drew attention immediately: it promises, at least for men, a simple and uncontroversial method for harvesting therapeutic stem cells. Two other research groups have made similar claims in the past, and a patent showdown among the three is looming.

As Hasenfuss's team reports in *Nature*, the reprogrammable cells come from sperm-producing stem cells in mice (K. Guan *et al.* *Nature* doi:10.1038/nature04697; 2006). It is not clear whether the cells in the testes have the ability to produce all other kinds of bodily cells, as embryonic stem cells can. But after the team extracted, cultured and multiplied the cells 700-fold, they converted to an embryo-like state in 4 out of 15 cases. When injected into early embryos, the converted cells helped build various organs of the resulting mouse. And, *in vitro*, the converted cells turned into several different types of cell, including heart, brain and skin cells.

Hoping to move the work into humans, the team has already received tissue samples from testicular-cancer patients. Hasenfuss thinks that, eventually, a simple biopsy could extract the necessary cells. Such cells would be a genetic



Hopes and fears: sperm-producing cells can stand in for embryonic stem cells, in mice at least.

match with the patient, which makes them less likely to be rejected if used in therapy.

But Hasenfuss's work has its sceptics, including Takashi Shinohara, a mouse germ-cell expert at Japan's Kyoto University. Shinohara's group was the first to isolate reprogrammable cells from mouse testes, but it used two-day-old pups (M. Kanatsu-Shinohara *et al.* *Cell* **119**, 1001–1012; 2004). His experience differs from that of the Göttingen researchers in some ways — for example, they report that stem cells injected into embryos are able to integrate into various tissues. “Perhaps they have some dif-

ferent kind,” he says, “but I don't think that type of cell exists.”

Shinohara also points out that many researchers have tried doing similar experiments, with more sophisticated culture media, without the same results. “It's too good to be true,” he says. Kaomei Guan, of the Göttingen team, counters that simplicity may be the key. Applying a complicated mix of certain growth factors in a lab dish for a long period of time, he says, can cause other cells to take over and the sperm stem cells to die.

A third, US-based group has already claimed

Space scientists get double reprieve

WASHINGTON DC

NASA reversed two decisions in one day on 27 March, reinstating the Dawn asteroid mission cancelled earlier this month and restoring some funds cut from astrobiology research. Both announcements have left scientists happy but also puzzled by a seemingly erratic decision-making process at the space agency, where managers are struggling to pay for an ambitious slate of science missions with a shrinking budget.

The US\$446-million Dawn

mission is expected to launch in July 2007 to study the asteroids Vesta and Ceres. The mission had been scheduled to launch this June, but last autumn, NASA stopped work on the project, owing to technical problems and cost overruns.

Although an independent assessment team reported in January that there were no technical barriers to launch, NASA associate administrator Mary Cleave cancelled the mission on 2 March. Her decision angered space scientists in the

United States and in Germany, which is contributing funds and expertise to the mission. Charles Elachi, director of the Jet Propulsion Laboratory in Pasadena, California, where Dawn is managed, asked NASA headquarters for a review. A team headed by associate administrator Rex Geveden decided on 23 March that the technical and cost issues were sufficiently in hand to permit launch.

Scientists attending an astrobiology meeting from 26 to 30 March in Washington DC also heard

of a reprieve, albeit a partial one. Carl Pilcher, NASA's senior scientist for astrobiology, had informed the community just last week that projected budget cuts of 50% meant that any grant proposals received in 2005 would be unlikely to be funded. He also said there would be no solicitation for proposals this year. At the Washington meeting, Pilcher said NASA headquarters has decided to restore enough money to award at least half the expected number of grants from 2005. Most

**STEM CELLS IN FOCUS**

Find a collection of News on this topic online.

www.nature.com/news/infocus/stemcells.html

some success in related experiments with human cells. Francisco Silva is vice-president of research at PrimeGen Biotech, based in Irvine, California. He says that, working with samples from 22 testes, the company “has successfully therapeutically reprogrammed human germ cells and differentiated them to multiple cell types *in vitro*”. Silva has presented the data at scientific meetings, and says he plans to submit results for publication soon. But Shinohara is again sceptical.

Hasenfuss, Shinohara and PrimeGen all say they have filed for international patents on their technologies. Some legal squabbles are likely. Hasenfuss, for example, claims work in adult mice is different from that in neonatal mice. Shinohara says the two experiments are essentially the same.

Whether any of the mouse-based claims would cover humans is not clear. Even if a patent office grants a patent, it may be invalidated by the courts, says Lynn Pashow, an intellectual-property attorney with Fenwick & West in Mountain View, California (see ‘US to rule on research patent’). “It may be difficult to know the scope of the valid claims,” he says.

In any case, researchers still have to show that the cells can be derived, grown and manipulated just like those from the alternative source, cloned embryos. In theory, the reprogrammable testes cells could circumvent the ethical difficulties of stem-cell work that involves destroying human embryos. “This could put the embryonic stem-cell people out of business,” says Peter Donovan, a stem-cell expert at the University of California, Irvine. “But it remains to be seen whether they work in humans.” ■

David Cyranoski



NASA's Dawn asteroid mission has been revived.

astrobiologists at the meeting applauded the news, but vowed to keep lobbying for more money. “We can see this as the beginning of a negotiation,” says Baruch Blumberg, a Nobel laureate and former director of the NASA Astrobiology Institute. ■

Tony Reichhardt

US to rule on research patent

WASHINGTON DC

It is a case that questions the very nature of what can be patented. Before the US Supreme Court adjourns at the end of June, it may decide whether a patent based on a biological relationship between two substances can be issued. A verdict could have implications for many other US and worldwide patents.

“The United States is often an incubator for these issues,” says Francis Gurry, deputy director-general of the World Intellectual Property Organization in Geneva, Switzerland. “The world is watching with a great deal of interest.”

A natural phenomenon or a law of nature generally cannot be patented, but a process that takes advantage of that phenomenon or law can be. Lawyers have been battling over this rule in biotechnology cases since at least 1980, when the Supreme Court ruled that a newly discovered microbe capable of digesting petrol could be patented. That case opened the door for the biotechnology industry and ushered in patents for antibodies, microorganisms and cells.

Some feel that it is now too easy to patent natural phenomena. “There might be some individuals who would like to close that door,” says Nick Godici, a former US Patent and Trademark Office commissioner.

The latest case — Laboratory Corporation of America Holdings (LabCorp) against Metabolite Laboratories — deals with the field of biomedical diagnostics. It stems from a patent-infringement case over



Thought police: the Supreme Court is reviewing a case that could “wreak havoc on the patent world”.

a method for diagnosing vitamin B deficiencies. In 1990, Metabolite patented a diagnostic test, based on an assay that measures blood levels of homocysteine, an amino acid. High levels of homocysteine are correlated with low levels of vitamin B₁₂. Physicians often order such tests because high homocysteine levels are also correlated with increased health risks such as heart attacks, stroke and birth defects.

Metabolite licensed the test to LabCorp, a clinical testing company. LabCorp stopped using it in 1998 and replaced it with a similar test developed by another company. When LabCorp stopped paying royalties, Metabolite sued for patent infringement. LabCorp lost, was ordered to pay \$7.8 million, and lost again on appeal.

In the case heard by the Supreme Court on 21 March, LabCorp argued that Metabolite had patented a law of nature by asserting ownership of the relationship between homocysteine levels and vitamin B₁₂. Under these circumstances, LabCorp said, physicians infringe the patent simply by thinking about the relationship when studying test results.

Miguel Estrada, an attorney for Metabolite, told the court that the natural relationship is integral to the diagnostic step. Addressing the broader issue, he argued, could “wreak havoc on the patent world”.

A corresponding European patent does not claim to cover the relationship between levels of homocysteine and vitamin B₁₂, and is not currently being challenged, says Siobhan Yeats, a biotechnology director at the European Patent Office in Munich, Germany.

The Supreme Court could now come to a verdict, pass the case back to the lower courts, or dismiss it altogether. One complication is that the biological-relationship argument was not addressed in the lower courts.

And although no one can predict which way the vote will go, at least one justice indicated his thoughts during the hearing. Justice Stephen Breyer hinted that patenting a scientific phenomenon might limit researchers’ motivations to search for new cures. “If you don’t provide them with an incentive,” he said, “they may think of less.” ■

Jacqueline Ruttimann

W. K. HARTMANN/UCLA

ON THE RECORD

“Nobody cares about contaminating the Moon. It’s like New Jersey.”

Astrobiologist Chris McKay explains why early exploration missions will test technologies on the Moon, rather than on Mars.

“I’m kind of disappointed, but I guess I did OK.”

Fifteen-year-old Gaurav Rajav fails to break a North American record for reciting the digits of pi. He masters 8,784 digits.

Sources: *News@nature*; AP

SCORECARD**Stuffed animals**

The yeti crab, recently discovered in the South Pacific deeps, is honoured with its own sewing pattern — letting you recreate it as a plush toy.

**Chopsticks**

The Chinese government adds a 5% tax to disposable wooden chopsticks, in an attempt to slow deforestation.

**Organs**

Europe restricts the building of pipe organs, saying the instruments violate a European Union directive to limit the amount of lead contained in products that work on electricity.

NUMBER CRUNCH

Nobel-prizewinning physicist Carl Wieman announced last week that he is leaving the University of Colorado, where he has worked for 22 years, and joining the University of British Columbia. Why?

\$227,000 is Wieman’s base salary at Colorado.

\$3 million is the amount paid to Colorado’s football coach after he was fired following several scandal-laden seasons.

\$10 million is the amount the Canadian university has promised Wieman for his science-education efforts. Wieman is taking a salary cut to go there, but says that at least British Columbia values academics as highly as athletes.

Sources: *Rocky Mountain News*, *Colorado Daily*

Scans suggest IQ scores reflect brain structure

Researchers say that a remarkable data set on the developing brain adds to the idea that IQ is a meaningful concept in neuroscience. The study, which is published on page 676 of this issue, suggests that performance in IQ tests is associated with changes in the brain during adolescence.

Claims that IQ is a valid measure of intelligence tend to attract angry responses, in part because of studies that have attempted to link group differences in IQ with race. In their 1994 book *The Bell Curve*, political scientist Charles Murray and psychologist Richard Herrnstein argued that the lower-income status of some US ethnic minorities was linked to below-average IQ scores among those groups. These were in turn attributed to mainly genetic factors.

Before that, Harvard University entomologist Edward Wilson provoked outrage with work that proposed evolutionary explanations for human behaviour and individual differences in intelligence; critics called the work racist. And this month, the journal *Intelligence* printed an editorial note defending its policy regarding the publication of controversial papers. The note comes after a study linking IQ and skin colour (D. I. Templer and H. Arikawa *Intelligence* 34, 121–139; 2006), published online last November, prompted a string of complaints from scientists.

Yet researchers studying IQ say the social climate is becoming more receptive to such studies, in part because it is now widely agreed that cognitive abilities are shaped by environmental factors as well as genetic ones.

The latest result, from a team led by Philip Shaw at the National Institute of Mental Health in Bethesda, Maryland, adds to the debate by linking IQ with changes in the brain over time, rather than fixed attributes such as brain size. “It’s not that brainy children have more grey matter,” says Shaw. “The story of intelligence is in the trajectory of brain development.”

Shaw’s team tracked a group of more than 300 children as they aged from 6 to 19, running them through a series of cognitive tests — IQ is determined by combining scores from tests of a range of verbal and non-verbal abilities. The team also measured the size of brain structures using magnetic resonance imaging at roughly two-year intervals: more than half the children had at least two scans, and around a third were scanned three or more times.

When the researchers split the children into three groups according to their initial IQ scores, they noticed a characteristic pattern of changes in the brains of the group with the highest scores. The thickness of the cortex — the outer layer of

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

Italians put science chief on the spot

Italian scientists are pressuring the president of the National Research Council (CNR) to justify his plans for reforming the agency — and to come clean about his academic background.

Up to 1,000 scientists from across Italy are expected to gather in Rome on 30 March to demonstrate over their unhappiness with the CNR’s management team — installed by Prime Minister Silvio Berlusconi’s government two years ago. The council runs almost 100 of Italy’s research institutes.

“We are particularly concerned about the administrative council’s plans to restructure the CNR’s institutes, merging them into 67 entities — based only on bureaucratic not scientific considerations,” says Rino Falcone of the CNR Institute of Cognitive Sciences and Technology in Rome.

Falcone helped to organize the march following a report in *Nature* earlier this month that described researchers’ increasingly public dissatisfaction with the government’s science policy (see

W.S.D. MCINTYRE/CORBIS

the brain that controls high-level functions such as memory — started off thinner than that of the other groups, but rapidly gained depth until it was thicker than normal during the early teens. All three groups converged, with the children having cortexes of roughly equal thickness by age 19. The strongest effect was seen in the prefrontal cortex, which controls planning and reasoning.

“My first impression was ‘wow, this is amazing,’” says Jeremy Gray, a psychologist at Yale University in New Haven, Connecticut. He notes that it is difficult to persuade children and parents to return for scans over a long period of time, so imaging studies are usually limited to tens, rather than hundreds, of subjects.

Shaw’s study raises questions that could prompt further research. His team did not look

at what could be causing the changes in cortical thickness, for example; the group points out that several mechanisms — including the formation and elimination of connections between brain cells — could be responsible. Also unknown is how genetic and environmental factors contribute to the change.

The study is likely to prompt discussion of the possible social applications of such results, but these are limited. The trend identified by Shaw was significant when results from all the subjects were combined, but would probably be too small to predict how an individual child is likely to fare in school, for example.

There are also likely to be queries about whether the research should have been conducted in the first place. IQ is a good predictor of performance at school and in the workplace. For some neuroscientists, this makes the physiological factors that contribute to IQ worth studying, in order to probe how intelligence works. “There’s good evidence from functional imaging studies that very demanding tasks activate the prefrontal cortex, and that activity correlates with IQ,” says Shaw. “We’re getting at some common processing resource.”

Gray also points out that metrics related to IQ can help predict speed of recovery from stroke, so studying them could lead to new therapies. But for many others, the concept of linking IQ and intellect remains socially dangerous and scientifically dubious. Steven Rose, a neuroscientist at the Open University in Milton Keynes, UK, says performance on cognitive tasks depends on a large range of factors, from emotive state to recall ability, and that the “IQ approach ignores all of these”. He adds that even as a predictor of ability in school it has traditionally been put to negative ends — to weed out, rather than help, less able children. “We shouldn’t go back to measures developed in the 1900s,” he says. ■

Jim Giles

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Children who perform differently
in IQ tests show different
patterns of brain development.

Nature 440, 264–265; 2006).

The *Nature* article also reported that CNR president Fabio Pistella had supported his candidature for the presidency with a CV claiming 150 publications. Only three have been identified by Thomson ISI searches, but Pistella retorted that his role required only “management skills”. On 21 March, 39 of Italy’s most highly cited scientists sent an open letter to Pistella claiming that this response to the “serious” charge “did not fit the dignity” of his function as CNR president.

The letter also accuses Pistella of having helped cause the decline of Italy’s energy and environment agency — which he

headed before joining the CNR. It asks him to clarify the criteria for his proposed restructuring of the CNR so as to prevent it from suffering the same fate as the energy agency.

Finally, the letter challenges Pistella to reveal the details of his scientific publications. “We believe this is the only way to restore a relationship of trust and dialogue with the entire national and international scientific community,” it says. Pistella has responded with a promise to publish his full reference list on the CNR website. But he had not done so by the time *Nature* went to press. ■

Alison Abbott

SPECIAL REPORT

Credit where credit's due

Disputes over who truly deserves the credit — or blame — for published work can result in bruised egos, damaged careers and court cases. **Helen Pearson** looks at ways to avoid fights over authorship.

A flurry of squabbles about high-profile biological research is prompting scientists to revisit a perennially touchy subject: how should credit for scientific findings be assigned?

In recent months, a panel on research integrity at the University of Pittsburgh, Pennsylvania, scolded cloning expert Gerald Schatten for his limited contribution to papers he co-authored with a South Korean team. Ian Wilmut of the Roslin Institute in Edinburgh has been criticized for taking most of the credit in 1997 for the cloning of Dolly the sheep. And a co-author has accused Alison Murdoch, of the University of Newcastle upon Tyne, UK, of hogging the credit for an advance in cloning human embryos (see 'Cloning clashes').

These quarrels all involve cloning, but they

resonate with a wider audience. Scientists often work for years towards a breakthrough; the recognition for it can bring prestige, promotion and pay. Small wonder, then, that they care about how credit is assigned.

"This is about one of the most important questions you can ask, because credit more than anything drives science," says Drummond Rennie, deputy editor of the *Journal of the American Medical Association (JAMA)*. But he and others say that there are some simple courtesies and practices that could, if widely adopted, help ease any tension.

Most concern focuses on who gets listed as an author on a publication. Tallies of such publications are one of the main measures of scientific achievement. And it is particularly important for a key finding: "The difference in

authorship at the time can end up being the difference between a Nobel Prize and not," says Nicholas Steneck, an expert in research ethics at the University of Michigan in Ann Arbor.

By some measures, abuses of and disputes over credit are commonplace. In 2005, a team of researchers from Minnesota found that 10% of US scientists surveyed admitted inappropriately assigning authorship in a paper¹. And in a study of young physicists, some 4% reported that they had seen cases in which relevant studies were not cited in a paper². The practice infuriates many scientists, because they feel they are being cheated of acknowledgement, and because citation counts are another measure of a paper's — and an author's — worth.

The central problem is that there are no set

CLONING CLASHES

Several prominent cloning researchers have recently found themselves in disputes over who should take credit for major discoveries.

- At an employment tribunal this month in Edinburgh, UK, Ian Wilmut of the Roslin Institute said that he played a mostly supervisory role in the cloning experiments that spawned Dolly the sheep³, and that most of the credit should go to co-author Keith Campbell. The

admission came as little surprise to many in biomedical science: a laboratory leader's main contribution is often ideas and guidance rather than experiments. But bitterness remains: one of Wilmut's technicians, Bill Ritchie, told the *Guardian* newspaper that

more of the credit should have gone to him. When approached by *Nature*, Wilmut said: "The legal case is still going on and I am afraid that in these circumstances I cannot comment."

- In January, stem-cell biologist Miodrag Stojkovic (left) went public with claims that he left the University of Newcastle upon Tyne, UK, partly because his colleague Alison Murdoch (far left) took public credit for controversial human nuclear-transfer experiments⁴. The university says that Stojkovic was consulted on all publicity relating to his work, and that he received credit in many major media outlets.

- Last month, the University of Pittsburgh in Pennsylvania chastized its biologist Gerald Schatten (right) for assuming co-authorship on two papers⁵⁻⁷ from the lab of disgraced stem-cell scientist Woo Suk Hwang of Seoul National University in South Korea.



Y.-J. JUNG/AP

He was quick to distance himself from the work when doubts were raised about the veracity of the data. A report from a university investigative panel was particularly biting about Schatten's contribution to the report of the cloned dog Snuppy. It said that "his major contribution to the paper was a suggestion that a professional photographer be engaged so that Snuppy would appear with greater visual appeal. It is less clear that this contribution fully justifies co-authorship." **H.P.**



NORTH NEWS AND PICTURES



N. FEANNY/CORBIS SABA

Who cloned Dolly? The work brought fame for Ian Wilmut, above; he says a colleague deserves most of the credit; and a lab technician feels unfairly left out.

rules for assigning credit. Biologists tend to place a supervisor or lab head last in an author list; organic chemists might put him or her first. Until recently, it was standard for the head of a German department or institution to take credit on a paper regardless of input. Graduate students and postdoctoral fellows frequently complain of being pushed down the author list, if they are included at all.

These differences in expectation create fertile ground for disputes. And many say that the number and ferocity of such clashes is rising with the number of collaborative projects, which often have multiple authors spanning disciplines and national borders. In genome sequencing and particle-physics collaborations, for example, a paper's author list can run into the hundreds. Some of these fights become public when there is disagreement about who should talk to, and be acknowledged in, the press. "Some people are very good at co-opting the media and blowing their own horn," says Martin Blume, editor-in-chief of the American Physical Society.

Many institutions, societies and journals have guidelines on what deeds warrant a name on a paper. According to the International Committee of Medical Journal Editors (ICMJE), authors must have made a substantial intellectual contribution to a study's conception and design, or to the acquisition,

analysis or interpretation of data. They must also have drafted or revised the article's intellectual content, and approved the final version.

Ground rules

On this basis, a technician who performed routine genetic sequencing, a colleague who supplied an antibody, or a departmental bigwig who supplied funding would be relegated to the acknowledgements. The ICMJE guidelines also say that authors should be able to take public responsibility for appropriate portions of the work — in other words, they must take the flak if the results come under fire, as Pittsburgh's Schatten found.

Whatever the rules, many scientists feel that the reality is different. "We all know labs where powerful figures appear on papers that are not within their field of expertise," says Ray Dolan, who heads a neuroscience imaging department at University College London. A particular problem, he says, is researchers who invent a technique and then demand credit every time it is used, even after it becomes standard procedure.

One way to avoid ambiguity is for each author to spell out their contribution in the paper. Many medical journals, including *JAMA*, require this. The practice is voluntary at *Nature*; editors at *Science* are considering whether to introduce it. "It's moral, it's easy

and it takes up virtually no space," says Rennie, who has pushed for adoption of the practice.

Training scientists in the ethics of authorship could also help; many such courses already exist. In his ethics course, neuroscientist Michael Zigmond, of the University of Pittsburgh, asks graduate students and postdocs to consider a case study of ten people and reach a consensus on whose contribution merits authorship. The participants, he says, often disagree at first, but eventually agree on criteria.

Indeed, many scientists say that the simplest way to avoid fights is to discuss authorship when a collaboration begins. "Proactive discussion is really important," says Kate Kirby of the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts, who co-authored the survey of young physicists.

The stakes are perhaps highest for the most prestigious scientific prize, the Nobels. Goran Hansson, who chairs the Nobel Committee for Physiology and Medicine, says it would help his task if the research community could agree on criteria for co-authorship. "We would still have to be very meticulous," he says, "but it would be a bit easier."

"Some people are very good at co-opting the media and blowing their own horn."
— Martin Blume

1. Martinson, B. C., Anderson, M. S. & de Vries, R. *Nature* **435**, 737–738 (2005).
2. Kirby, K. & Houle, F. A. *Phys. Today* **57**, 42–46 (2004).
3. Wilmut, I., Schnieke, A. E., McWhir, J., Kind, A. J. & Campbell, K. H. S. *Nature* **385**, 810–813 (1997).
4. Stojkovic, M. et al. *Reprod. Biomed. Online* **11**, 226–231 (2005).
5. Hwang, W. S. et al. *Science* **308**, 1777–1783 (2005).
6. *Science* **311**, 335 (2006).
7. Lee, B. C. et al. *Nature* **436**, 604 (2005).



BRIGHT FUTURE FOR SUN'S TWIN

Homely stars make perfect targets for planet-finders.

www.nature.com/news

More whale strandings are linked to sonar

Examinations of four whales found stranded along the Spanish coast in January seem to confirm a 2003 *Nature* report linking sonar to the deaths of several beaked whales.

In recent years, naval sonar devices have been the suspected cause of an increasing number of whale strandings worldwide. The whales are thought to take evasive action to avoid the noise, sometimes diving and surfacing until they suffer decompression sickness and die.

In 2003, British and Spanish researchers reported that Cuvier's beaked whales (*Ziphius cavirostris*), stranded off the Canary Islands the previous year, had deadly gas-bubble lesions called emboli in their livers. They suggested these were caused by decompression (P. D. Jepson *et al.*



Air bubbles have been found in the tissues of dead whales discovered in Spain.

Nature 425, 575–576; 2003).

After a group of beaked whales went ashore in January, along Spain's Costa del Sol, the Spanish Cetacean Society in Madrid called veterinarian Antonio Fernandez to perform necropsies on four of the animals. He and his colleagues from the University of Las Palmas de Gran Canaria found the same

embolic syndrome as that found in the 2003 study.

"This is the first confirmation of the 2003 report," says veterinarian Paul Jepson of the Zoological Society of London, lead author of that article. The new findings are expected to be published in coming months.

Officials at the Cetacean Society

suspect that mid-frequency naval sonar caused the strandings. But Fernandez notes that the ships that might have been responsible have not been identified.

Earlier this month, about 45 pilot whales died after stranding on the western side of the island of Sulawesi in Indonesia, following joint US and Indonesian naval exercises in the nearby Macassar Strait. The cause of the stranding is under investigation.

Some US Navy officials, and oceanographers who use devices to generate air bursts underwater for seismic studies, have been accused of blocking efforts to uncover the links between noise and whale strandings (see *Nature* 439, 376–377; 2006).

Rex Dalton

A. FERNANDEZ

Chemistry lab blast kills French researcher

A massive explosion devastated a three-storey chemical research laboratory in northeastern France on 24 March, killing one researcher and seriously injuring a student. The accident occurred on the campus of Upper Alsace University, in a lab that dealt with the analysis and prevention of industrial chemical accidents.

Investigators have yet to determine the exact cause of the blast. One hypothesis is that a cylinder of ethylene leaked the odourless gas, which is highly explosive

when mixed with air at high concentrations. The scientist killed was Dominique Burget, 41, a photochemist, who was in his office on the floor above the blast.

New York universities get \$400-million hand-out

New York City intellectuals received a boost last week from two \$200-million donations, among the largest for US academia in the past decade. One donation will establish a neuroscience centre at Columbia University; the other is for a New York University (NYU) institution to study the ancient world.

Columbia's gift, from the Jerome L. Greene Foundation, is the largest in the university's history. Researchers at the centre will tackle all aspects of brain science to try to understand how the molecular workings of brain cells drive human behaviour and diseases such as Parkinson's and Alzheimer's. It will be led by three eminent Columbia neuroscientists: Thomas Jessell, Richard Axel and Eric Kandel.

NYU's Institute for the Study of the Ancient World, created with a gift from the Leon Levy Foundation, will support research and teaching in disciplines including archaeology, history, geography, economics and sociology.

Pakistan aims to build six science universities

Pakistan's government has announced an ambitious plan to build six new science and technology universities by 2008.

The universities would "go a long way in providing Pakistan with cutting-edge technology in various fields", Pakistan's President Pervez Musharraf told staff at the University of the Punjab in Lahore on 13 March. He said that the universities would be partly funded with aid from five European nations and South Korea.

The plan is part of a recent revival of Pakistan's higher-education system. Over the past three years, the country's higher-education budget has risen by 1,500% to US\$190 million, says Atta-ur-Rahman, who chairs the higher-education commission.

Lower fluoride limits to save kids' teeth, urges US study

The maximum allowable limit of fluoride in US drinking water is too high, says a report commissioned by the Environmental Protection Agency (EPA).

The EPA currently enforces a limit of 4 milligrams of fluoride per litre of water.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

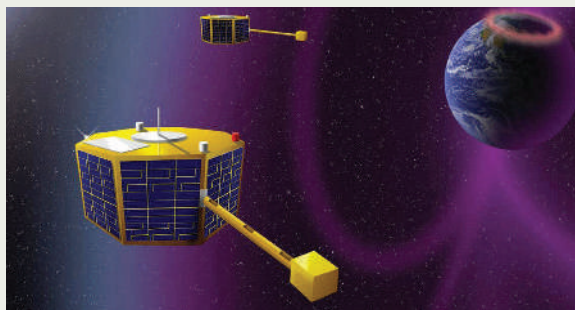
Firemen search through the wreckage caused by a fatal accident at Upper Alsace University.

V. KESSLER/REUTERS

NASA's tiny satellites fly, but private rocket crashes

NASA's Space Technology 5 mission — a trio of hatbox-sized satellites meant to test miniature components and techniques for formation flying in space — launched successfully on 22 March (artist's impression pictured). A Pegasus rocket fired from an airplane carried the satellites into polar orbit. Scientists hope that future swarms of mini-satellites will have many scientific applications, including studies of space weather from different vantage points in the inner Solar System.

Not so lucky was the privately funded Falcon 1 launcher built by Space Exploration Technologies of El Segundo, California, which



failed 30 seconds into its first flight on 24 March. Early analysis pointed to a fuel leak as the cause of an onboard fire that brought the rocket down less than 100 metres from its launch pad on the Pacific atoll of Kwajalein. The company's chief executive, Elon Musk, says he hopes to try again within six months.

research, and to support young researchers and women.

With a total five-year budget of up to ¥25 trillion (US\$215 billion), the plan will try to diversify investment that has so far concentrated on areas such as the life sciences and nanotechnology.

The plan was produced by the government's Council for Science and Technology Policy, which marked out 62 strategic projects as priorities for investment. Each will be assigned targets: for example, launching rockets more than 20 times with a 90% success rate, or becoming the world leader in stem-cell technologies.

"Giving practical targets is an easy way to make the public understand how taxpayers' money is used," says Yasuharu Suematsu of the National Institute of Informatics in Tokyo.

Corrections

In the Editorial "Everybody's fault" (*Nature* **440**, 1; 2006), the location of the proposed scientific peace park should have been identified as the hotly disputed Siachen Glacier. And in an accompanying News Feature on page 16 of the same issue, Jack Shroder was wrongly described as a seismologist; he is a geomorphologist.

Due to an editing error, our piece "Are adverts revealing nuclear secrets?" (*Nature* **440**, 389; 2006) talked of uranium-238 as the fissile isotope, when it should have referred to uranium-235.

But the report says that 10% of children in areas with that much fluoride suffer from severe enamel fluorosis, in which teeth become brown and pitted. Most on the panel agreed that this condition, which is found around the world, can be considered an 'adverse health effect'.

The 450-page study also notes that there are conflicting data on whether high levels of fluoride can be correlated with bone

cancers. The EPA says it will now re-evaluate its enforceable limits.

Punchy plan gives Japan's scientists explicit targets

On 1 April, Japan will begin its third five-year science and technology plan. This aims to achieve concrete results from

NASA

THE SON ALSO RISES

The two Roger Pielkes can be obstructionist pains in the neck, say their colleagues. So why is this likeable father-son pair such a welcome addition to the debate on global climate change? **Kendall Powell** clears the air.

Roger Pielke Senior and Roger Pielke Junior share a name, a profession and a reputation. Both are mathematics-trained history buffs. Both ski and play golf as part of their active Colorado lifestyles. And both are prominent scholars in the highly polarized field of climate science, where their name can provoke much eye-rolling.

The elder Pielke, 59, is professor of climatology at Colorado State University and the state's official climatologist. The younger Pielke, 37, is an expert in science policy at the University of Colorado, with a bumper sticker that declares 'Question Predictions' in his office. Father and son share a proclivity for contentious, if polite, debate, and they both antagonize their colleagues more often than their affable exteriors would suggest.

Yet there are notable differences. Pielke Sr is a true climate hound, steeped in decades of research on atmospheric science. By contrast, Pielke Jr is a self-described policy wonk, who claims he simply hasn't inherited his father's obsession with the weather.

Junior does, however, have the famous Pielke tenacity, and has put it to use in the world of science policy. He caught the bug after interning on Capitol Hill in 1991, when his adviser Rad Byerly became the chief of staff for the House Committee on Science. Pielke Jr then returned to the University of Colorado in Boulder to finish his master's degree, with a thesis that calculated the true cost of a space shuttle launch.

He concluded that each launch cost just over \$1 billion,

contrasting with NASA's estimate of \$400 million¹. Shortly after his numbers appeared in a 1993 article in *The New York Times*, Pielke Jr took a call from an official at NASA's Johnson Space Center, who asked him to retract his conclusions about the cost. He said he gladly would, if the official could only pinpoint what exactly was wrong. The person never called back.

The incident, says Byerly, demonstrates the younger Pielke's coolness under fire. "He knows right where the jugular is," says Byerly.

For his doctorate work, Pielke Jr turned to the stickiest problem he could think of. "I asked myself: what's the hardest possible evaluation problem that I could do, that's messy and involves politics?" In the early 1990s, the obvious choice was climate-change policy. And so he rigorously evaluated the US Global Climate Research Program, concluding that it was not meeting its mandate of providing useful information about climate science for decision-makers².

From that thesis arose an idea that Pielke Jr continues to push today, much to the discomfort of some climate scientists. He argues that the traditional relationship between science and policy, in which scientists do good science and hand the results to the policy-makers, is obsolete — particularly for complex modern issues such as stem-cell research and climate change. He advocates a two-way approach, in which policy-makers point scientists at the next set of questions to which answers would be useful.

In the example of climate change, Pielke Jr says, many

FROM THE ATMOSPHERE TO THE BLOGOSPHERE

Roger Pielke Junior and Senior each run a widely read climate weblog. Here they tell *Nature* how blogging enhances their research.

ROGER PIELKE JR

Prometheus: The Science Policy Weblog
<http://sciencepolicy.colorado.edu/prometheus>



"It started as an experiment for our centre, and now it serves a number of different purposes. It is kind of like an extra hard drive for my brain. I can search for things that I've written, something I might want later, sort of like my professional notes in a public format."

"I'm surprised at the reach the blog has, which is rewarding for this centre with only eight of us here. We can put an argument on it and it shows up out there in the

real world. I get contacted by professionals in the United States or elsewhere that I would have never met otherwise."

"Blogs are also out there for the public, and it gives you an entirely different perspective on how well the public is getting your message."

"The blog is like an extra hard drive for my brain. I'm surprised at the reach it has."

ROGER PIELKE SR

Climate Science
<http://climatesci.atmos.colostate.edu>

"My weblog was completely motivated by my son's. I was sending all these e-mails out to people about committee reports and he said, 'Why don't you just do a weblog?'"

"With so many journals out there now, it is hard to keep track. When a peer-reviewed paper comes out, I can put up the abstract and a summary of key points on the blog."



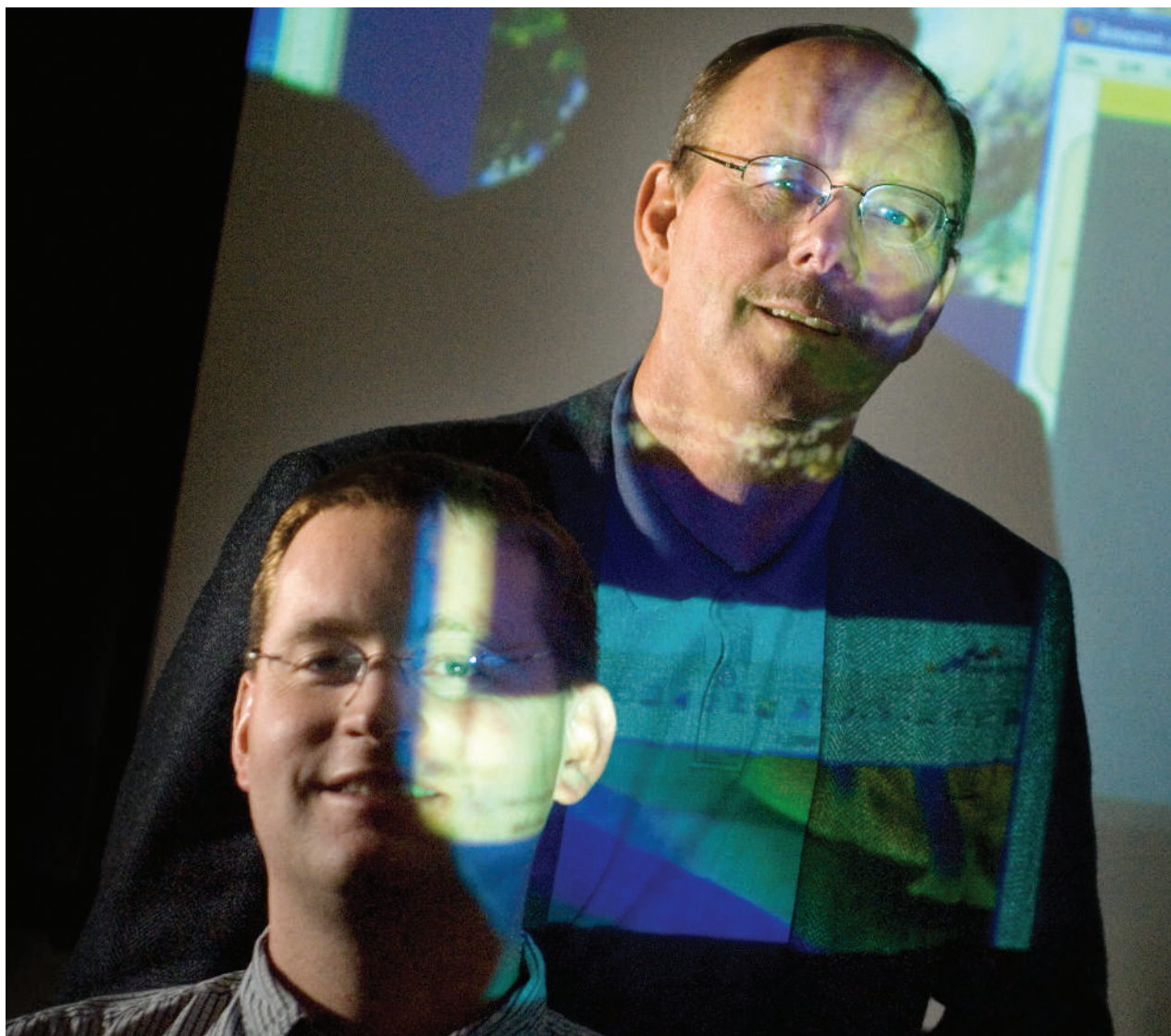
"Now I'm making my arguments to a broader community to see how well they stand up. I also use it as a professional diary and it has increased my network."

"The feedback has been wonderful."

"In science, you should come up with ways to resolve a conflict, not ignore it."

— Pielke Senior

Look out for the next generation: the Pielkes study different aspects of climate change with a similar intensity.



researchers have taken one of two sides: backing either mitigation policies to reduce greenhouse-gas emissions, or adaptation policies to deal with climate change as it occurs. "One of the most important roles science can play is to invent new options and introduce them to decision-makers," he says. "When scientists take sides, they are giving up that role." He persistently challenges scientists who he thinks are acting as advocates for a particular position, including members of the Intergovernmental Panel on Climate Change and scientists who run a blog called RealClimate.

"To be frank, that irritates the hell out of me," says Gavin Schmidt, co-founder of the RealClimate site and a climate researcher at NASA's Goddard Institute for Space Studies in New York. "What he considers to be advocacy, to me, that's just interacting in the public realm." Schmidt and Pielke Jr have never met in person, but have had heated exchanges in the world of blogs (see 'From the atmosphere to the blogosphere').

Winds of change

While the younger Pielke ruffles feathers in the climate community, his father has been fighting the same battle on a different front. Pielke Sr studied the human impacts on climate long before it was a trendy field.

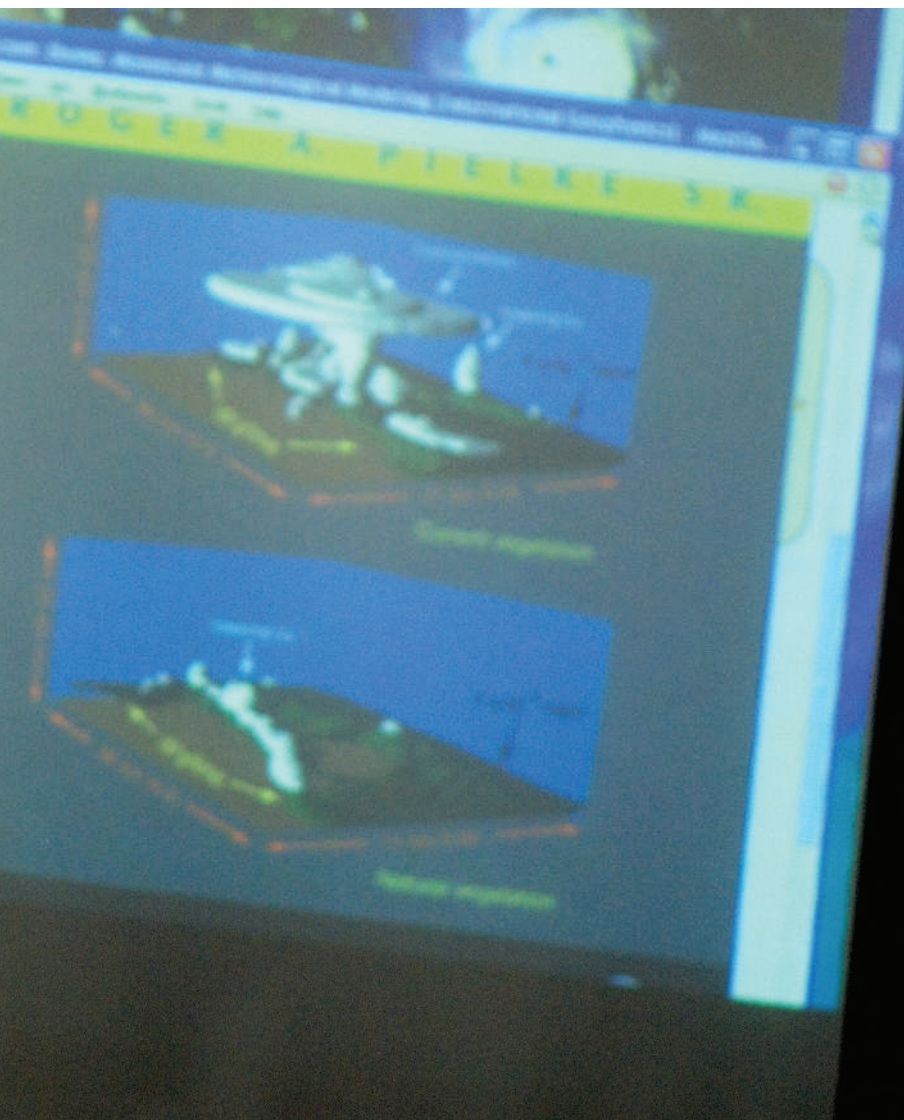
His PhD thesis, in the early 1970s, dealt with three-dimensional modelling of the Florida sea breeze and its

effects on severe thunderstorms. By the mid-1970s he was a professor at the University of Virginia, where he added the wetlands of the Everglades to his climate model and asked what would happen to regional climate if they were paved over³. By 1981 he had moved to Colorado State University in Fort Collins, where his group made breakthroughs in mesoscale atmospheric modelling, helping to develop the Regional Atmospheric Modeling System⁴.

Over the years, his research has looked at the effects of air pollution, aerosols and land-use changes on regional climate. It hasn't always been an easy ride. Last August, he resigned from a committee sponsored by the Climate Change Science Program (CCSP), which was preparing a report about temperature trends on Earth's surface and in its lower atmosphere, known as the troposphere.

Pielke Sr argued that members of the CCSP committee were focusing on their own work too much, and not including other perspectives that could explain possible discrepancies in the observed temperatures⁵. "If there is a disagreement in a science issue, you should come up with ways to resolve it, not ignore it," he says. And when he felt that a reporter on *The New York Times* had not accurately described his reasons for resigning, he launched an open letter on his blog to make his opinions known.

Others say the row was about more than including a variety of perspectives. They say it reflects part of a larger



tension in the climate community between those who do regional modelling, such as Pielke Sr, and those who work with global circulation models, which predict the planet's temperature for years to come. "The average global surface temperature is almost useless for what people care about — their growing season and where they live," says Pielke Sr. He argues that regional climate models that include climate forcings other than greenhouse gases, such as land-use changes, provide more useful information than the commonly used global circulation models.

Local heroes

In fact, neither father nor son thinks that predicting global average climate trends is possible or useful. Pielke Sr says that evaluating the sensitivities of local resources to climate change would be wiser — giving an idea of its effect on energy, water and the ability to respond to natural disasters, for example. Pielke Jr points out it that doesn't take precise climate predictions to begin assessing societal and economic vulnerabilities to climate change.

This may sound like common sense. But by questioning the global predictions that many climate scientists hold dear, the Pielkes often get mislabelled as climate sceptics. Their persistence inflames people's emotions, but it also wins them praise — sometimes from the same people.

Kerry Emanuel, a hurricane researcher at the Massachu-

"My father taught me how to disagree without being disagreeable."
— Pielke Junior

K. MOLONEY

setts Institute of Technology, has worked with Pielke Jr. "I think pushing people is a very laudable aspect of what he is doing — it helps focus on the truth. We, as scientists collectively, have become rusty on that." Emanuel and Pielke Jr have both underscored a little-publicized point in Emanuel's recent *Nature* paper on how hurricane intensity increases with increasing sea surface temperature — that the link does not explain the unprecedented damage from Hurricane Katrina^{6,7}.

Even Schmidt of RealClimate admits he has learned some lessons from Pielke Jr about how science gets misused in policy discussions. "He hasn't been afraid to interact with scientists," Schmidt says. "That interaction has not always been pretty, but he gets some kudos."

Professional mixers

Colleagues of the elder Pielke see similar merit behind his prodding of the CCSP committee and others. Dev Niyogi, Indiana's state climatologist at Purdue University in West Lafayette, says debate helps the climate community. "Many streams of thought are being constrained for political correctness, and science may not benefit from lack of discussion," he says. "We need someone to stir up the whole thing."

Pielke Jr credits his father with teaching him how to "disagree without being disagreeable". Asking hard questions is not always well received, he says. "But I learned how to be professional and respectful. One may be pushing against some cherished ideas or notions, but I think that's what makes science stronger."

Together, the Pielkes have become close professional colleagues, each approaching climate science from his own direction. Over the years, Pielke Sr says, he has come to better understand the interactions between scientists and policy-makers that his son promotes. And Pielke Jr has acquired unique access to climate researchers, becoming something of an 'embedded anthropologist' through his father's connections.

Father and son have published together, on such topics as the behaviour of hurricanes and their impact on society⁸. They rarely disagree on fundamental professional issues, but they do squabble over how much candy the grandchildren should consume. Gloria Pielke, wife of one and mother of the other, says that heated arguments are common, but never turn personal. "If someone disagrees," she says, "we just share our reasoning and then go on to the next hole of golf."

Friendly competition runs deep in this close family. Informal bets are common, from golf games to the weather. As the best golfer, Gloria is the safest bet on the greens. But when gambling turns to the weather, it isn't predictions from the Colorado state climatologist you want. It's those of his son, the policy wonk.

Kendall Powell is a freelance science writer based in Broomfield, Colorado.

1. Pielke, R. A. Jr *Aviation Week & Space Technology* 26 July, 57–58 (1993).
2. Pielke, R. A. Jr *Policy Sci.* **28**, 39–77 (1995).
3. Pielke, R. A. Sr et al. *Mon. Weath. Rev.* **127**, 1663–1673 (1999).
4. Pielke, R. A. et al. *Meteorol. Atmos. Phys.* **49**, 69–91 (1992).
5. Brumfiel, G. *Nature* **437**, 9 (2005).
6. Emanuel, K. *Nature* **436**, 686–688 (2005).
7. Pielke, R. A. Jr *Nature* **438**, doi:10.1038/nature04426 (2005).
8. Pielke, R. A. Jr & Pielke, R. A. Sr *Hurricanes: Their Nature and Impacts on Society* (Wiley, Chichester, 1997).



POWERHOUSE OF DISEASE

Many of the genes affecting mitochondria — tiny energy suppliers of cells — reside in the cell nucleus. **Nick Lane** joins the hunt for these sequences that may underpin diseases such as diabetes.

Some of Gerald Shulman's patients at Yale University School of Medicine are young and slim. There's little wrong with them, and probably won't be for a decade or two. Yet tests raise an ominous spectre. All are the children of parents with type 2 diabetes, and, already, in their twenties, they are becoming resistant to insulin, the hormone that should be keeping their blood sugar levels under control.

The problem seems to lie in their muscles, whose cells lack tiny lozenge-shaped structures called mitochondria. These normally function as powerhouses inside cells, burning up fuel with oxygen. Long regarded as the cell's menial coal-shovellers, mitochondria are emerging as key players in health and disease. The 'organelles' are unusual in having their own DNA, although many of the genes that once resided in the mitochondria have, over evolutionary time, decamped to the cell's nucleus. Shulman is one of a number of scientists who think that tracking down the hundreds of 'missing' genes that have shifted to the nucleus is going to change the way we think about common diseases such as diabetes and Parkinson's.

Mitochondria store the energy released from food in the form of a molecule called ATP, which is used to power virtually all forms of work in the body, from muscle contraction to protein synthesis. Your body's mitochondria generate an impressive total of some 65 kg of ATP every day. The double-membraned organelles (see picture, overleaf) perform this feat thanks to a process called chemiosmosis, which pumps protons across one of their membranes. ATP is generated when the current of electrically charged protons, produced by this pump, passes through tiny protein motors embedded in the same membrane.

Ancient union

As well as looking like them and using chemiosmosis in the same way as bacteria, mitochondria contain a bacteria-like genome. Indeed, mitochondria were once free-living bacteria; they were engulfed by larger cells two billion years ago in a unique merger that gave rise to all complex, or eukaryotic, cells. The size of the genome housed within the mitochondrion varies between species. All mammals, for example, have retained just 37 genes, whereas

yeasts have retained between 40 and 50, and some plants as many as 100.

But mitochondrial genomes did not start out so small — they probably once contained at least a few thousand genes, inherited from the free-living ancestor of mitochondria¹. Exactly what happened to most of these genes is a moot point, but the evolution of a stable symbiotic relationship within eukaryotic cells led to hundreds, perhaps even thousands, being simply transferred to the cell's main genome in its nucleus. These transfers meant that mitochondria became dependent on the host cell for virtually all their functions. Today, some 99% of human mitochondrial proteins are encoded in the nucleus; all the proteins and other molecules required to build mitochondria are synthesized in the main body of the cell, then imported into the organelle. Only a fraction of these genes has been identified; the rest lie hidden in the vast code of the nucleus's genome.

This enigmatic 99% is now the focus of intense scrutiny. There are good reasons to believe that genes affecting the mitochondria could play a central role in human health and disease. Most of the genes that have remained in the mitochondrion have been linked to a series of devastating diseases, indicating the importance of fully functional mitochondria to human health.

Genes residing in the mitochondria pose a particular problem, however — in part because

C. JAY



they are unusually prone to damage. Unlike nuclear genes, which are wrapped in protective proteins and stored safely away in the nucleus, mitochondrial genes are vulnerable to attack from highly reactive molecules called free radicals; these are generated during energy production. In mammals, the mutation rate of mitochondrial genes is 10 to 20 times higher than that of the nuclear genes.

The idea that mutations in mitochondrial DNA could cause metabolic diseases, or even ageing, has gained credence since Fred Sanger's group at the University of Cambridge, UK, sequenced the human mitochondrial genome² in 1981. According to David Thorburn, at the Murdoch Children's Research Institute in Melbourne, Australia, in the decades since, pathogenic mutations have been discovered in more than 30 of the 37 human mitochondrial genes. These alterations range from changes to single DNA bases to deletions of large sections of the genome. Their effects are a long list of rare disorders, best diagnosed and treated by specialists, who refer

to themselves as mitochondriacs.

The most common childhood condition is Leigh syndrome. This affects about 1 in 40,000 children and tends to develop within the first year of life, often after a viral infection. In most cases, degeneration of the central nervous system leads to loss of muscular coordination and death within a few years, although some children survive into their teens. Lethal infantile mitochondrial disease is much rarer but even more deadly. Children born after an uneventful pregnancy tend to have seizures soon after birth, make few or no spontaneous movements, and die of respiratory failure within weeks. Other conditions have relatively mild symptoms. A common feature of all these diseases is that they tend to worsen with age. Indeed, it is the cumulative effects of free-radical attacks, and the corresponding build up of mitochondrial mutations that may underpin aging.

Faulty engine

Mitochondria, along with their tiny genomes, are normally inherited only from the mother — they are present in huge numbers in the egg, whereas the handful in sperm is marked up for destruction in the fertilized egg. This gives at least some mitochondrial diseases a maternal-inheritance pattern. Even so, trying to spot mitochondrial diseases by looking to the mother can be grossly misleading, and has downplayed the importance of these organelles in disease. More than 80% of diseases known to be linked to faulty mitochondria don't follow a maternal-inheritance pattern at all.

Why not? At least partly because some mitochondrial diseases may be caused by mutations in the nuclear genes encoding mitochondrial proteins. So far, mutations in more than 30 nuclear genes have been shown to give rise to mitochondrial disease. Thorburn, however, estimates that as much as a tenth of the population may be carrying genetic disorders that could affect mitochondrial function³. This is based on estimates of the number of mitochondrial genes in the nuclear genome and the incidence of recessive genetic disorders. He echoes a favourite catchphrase of mitochondriacs: "Mitochondrial deficiency can theoretically give rise to any symptom, in any organ, at any age."

Other, more complex degenerative conditions, such as Parkinson's disease, progressive-blindness diseases and other nervous-system conditions also involve mutations in mitochondrial proteins⁴. Even cancer can be caused by mutations in nuclear genes encoding mitochondrial proteins⁵. Examples are now cropping up almost every year, and together they are beginning to focus attention on the central role of mitochondria in disease.

The actual contribution of nuclear genes to mitochondrial diseases is highly uncertain for a simple reason — we are surprisingly ignorant of what the nuclear genes actually are, and how they interact with mitochondrial genes. In mammalian mitochondria, the best guess is that the nuclear genome encodes 1,500 distinct mitochondrial proteins. So far, barely half have been formally identified, and of these, the function of a sizeable proportion remains unknown.

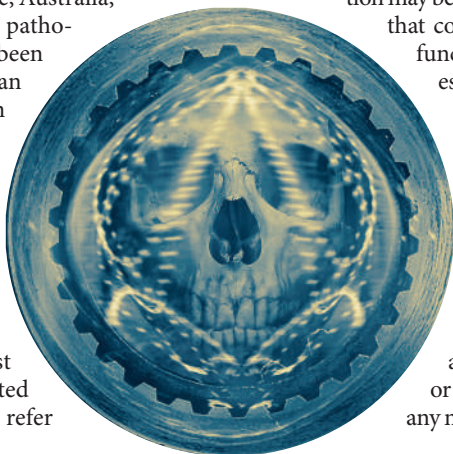
Nonetheless, the evidence that mitochondrial proteins are responsible for a lot more mischief than once thought is growing. A series of inherited conditions not thought of as 'mitochondrial' have turned out to be caused by mutations in genes encoding mitochondrial proteins⁴. For instance, Friedreich's ataxia (a progressive loss of coordination of voluntary movements) is caused by mutations in a gene encoding a small mitochondrial protein called frataxin. Hereditary spastic paraplegia (a progressive weakness and stiffness of the legs) can be caused by mutations in a mitochondrial enzyme, paraplegin.

Other, more complex degenerative conditions, such as Parkinson's disease, progressive-blindness diseases and other nervous-system conditions also involve mutations in mitochondrial proteins⁴. Even cancer can be caused by mutations in nuclear genes encoding mitochondrial proteins⁵. Examples are now cropping up almost every year, and together they are beginning to focus attention on the central role of mitochondria in disease.

These examples have all unexpectedly turned out to be 'mitochondrial', after years of tracking down candidate genes for the diseases. But new tools are letting scientists turn the old approach on its head. Rather than starting with an inherited condition and trying to track down the genes responsible, researchers are starting off with the mitochondria themselves, and attempting to hunt down the proteins needed to build them. Tracking down this array of proteins, or the mitochondrial 'proteome' is no easy task; researchers rely on a combination of methods to build an accurate picture, including mass spectrometry to identify proteins and molecular-biology techniques to measure RNA, the molecule used by cells as a template from which to build proteins.

All the techniques based on this bottom-up approach have strengths and weaknesses, but by taking the best information from each, scientists are gradually piecing the mitochondrial proteome together. Once the normal proteins have been identified, any oddities in patients can be

"Mitochondrial deficiency can theoretically give rise to any symptom, in any organ, at any age."



pinpointed. The abnormal protein can be mapped on to the candidate genes for disease, and any causal mutations involved identified.

In 2003, Vamsi Mootha, a computational biologist at the Broad Institute in Cambridge, Massachusetts, and his colleagues published a list of several hundred new mammalian mitochondrial proteins⁶, raising the known mammalian total to around 600. Crucially, however, Mootha's group also examined tissue variations. In mice, they found that around half the mitochondrial proteins identified were present in four different tissues — brain, heart, liver and kidney. But the other half tended to be tissue-specific, with some degree of overlap (around 50%) between different tissues.

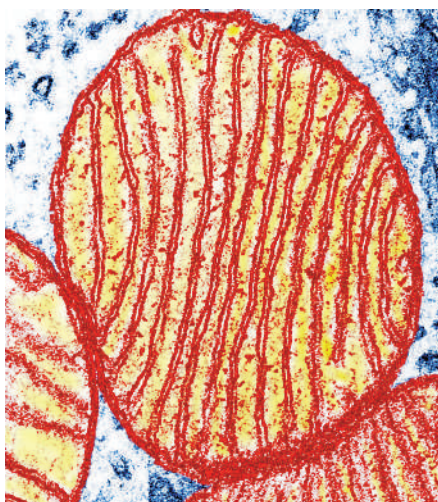
Building a powerhouse

Mitochondria are well known to carry out specific tasks in different tissues; for example, they make haem, part of the oxygen-carrying protein haemoglobin, in bone marrow cells. But the finding that hundreds of mitochondrial proteins varied in amounts from tissue to tissue came as a shock. If corroborated, this variation suggests that the control of mitochondrial gene activity is very sophisticated. And this has a corresponding impact on our susceptibility to disease; the more complicated the control system, the more likely it is to fail.

Mootha's group reported the first two tissue-specific mitochondrial proteins, known as *Errα* and *Gabpa/b*, in 2004 (ref. 7). Both control gene activity, which in turn affects how much mitochondria replicate themselves in particular tissues. If the expression of *Errα* and *Gabpa/b* is high, then mitochondria replicate at a high rate, and become densely packed in the tissue. If their expression is lower, the number of mitochondria and their ability to burn fuel falls. Critically, *Errα* and *Gabpa/b* influence mitochondrial function and density in particular tissues, notably the heart and muscle, and play a lesser role in tissues such as the liver. Mootha notes that this tissue specificity makes them valuable drug targets, because it restricts the potential for side effects in other tissues.

The next question for Mootha and his team was what happens if the activity of *Errα* and *Gabpa/b* falls? They predicted that a fall in the number and capabilities of mitochondria in particular tissues would result — a finding that Mootha and others had previously reported in the muscles of patients with diabetes. Sure enough, Mootha's lab found that the activity of these proteins was lower in the muscles of patients with type 2 diabetes⁸. But could such a change be a root cause of diabetes, or was this merely a consequence of some other metabolic problem, such as obesity?

Type 2 diabetes has two cardinal features.



Genes in the nucleus that encode proteins for the mitochondria (above) could underpin diseases.

The first is that cells become resistant to the effects of insulin, the hormone made by the pancreas that normally prompts them to take up and burn glucose. The second is high levels of glucose in the blood, or hyperglycaemia. Insulin resistance is typically one of the earliest signs of diabetes, often preceding hyperglycaemia by decades.

Faulty mitochondria have already been linked to the second phase of the disease — namely the emergence of hyperglycaemia. Defective mitochondria in the pancreas fail to burn sufficient glucose, so the levels of ATP in pancreatic cells are abnormally low. But these cells rely on ATP levels to help them estimate the amount of glucose in the blood. As a result, the cells do not sense glucose properly, do not release appropriate amounts of insulin and the blood glucose level creeps up⁹.

But what about insulin resistance? Shulman thinks that faulty muscle mitochondria could underlie insulin resistance in muscle tissue and was intrigued by Mootha's findings. "We've been working with volunteers who have a high genetic risk but a low 'lifestyle' risk of diabetes. We hope to eliminate confounding factors such as obesity, or indeed the early stages of diabetes itself, and focus on the earliest underlying genetic influences."

Complex pathways

Shulman's group has found three striking oddities in the muscle cells of the young volunteers: they are often very insulin resistant, taking up about 60% less glucose in response to insulin compared with the muscle cells of unaffected people; they have a low mitochondrial density,

about 40% lower than normal; and they have a large accumulation of fat molecules, or lipids, around 60% above normal¹⁰.

The key, says Shulman, is the high level of lipids. Lipids can cause insulin resistance by jamming the cellular machinery that helps receive the hormone's signal. But what causes their levels to rise in the cell? There are two main possibilities: a faster rate of lipid breakdown and delivery to muscles from fat tissues; or a defect in the muscle mitochondria themselves. If faulty mitochondria don't burn fats as fast as they should, then that could lead to a build-up of lipids inside the muscle cells. That would suggest the primary genetic cause of type 2 diabetes lies in the mitochondria. Faulty mitochondria also contribute to obesity, by not burning fats properly, and obesity in itself exacerbates diabetes.

Shulman's group could find no evidence that abnormal fat breakdown and delivery from fat tissues was responsible, and so turned to look at possible faults in mitochondria.

Following up on Mootha's findings, the team looked at whether a mutation in the genes controlling the tissue-specific mitochondrial proteins *Errα* and *Gabpa/b* could underpin the low density of mitochondria in the volunteers. The result, published in December last year, was a surprise. They could find no such mutations, implying that the reduction in gene expression measured by Mootha was not the primary cause of diabetes. The primary fault must lie in another, as yet unknown pathway governing mitochondrial proliferation and activity.

So faulty mitochondria may well be the cause of diabetes, but we still don't know what makes them faulty. Yet with hundreds of unknown mitochondrial proteins still to uncover, Shulman and Mootha have a long list of possible suspects to work through. Whether they will get results in time to help Shulman's young volunteers is an open question, but the answers seem set to revolutionize our understanding of disease.

Nick Lane is a science writer based in London.

"Faulty mitochondria may well be the cause of diabetes, but we still don't know what makes them faulty."

1. Timmis, J. N., Ayliffe, M. A., Huang, C. Y. & Martin, W. *Nature Rev. Genet.* **5**, 123–135 (2004).
2. Anderson, S. *et al. Nature* **290**, 457–465 (1981).
3. Thorburn, D. R. J. *Inherit. Metab. Dis.* **27**, 349–362 (2004).
4. Zeviani, M., Spinazzola, A. & Carelli, V. *Curr. Opin. Genet. Dev.* **13**, 262–270 (2003).
5. Gottlieb, E. & Tomlinson, I. P. M. *Nature Rev. Cancer* **5**, 857–866 (2005).
6. Mootha, V. K. *et al. Cell* **115**, 629–640 (2003).
7. Mootha, V. K. *et al. Proc. Natl Acad. Sci. USA* **101**, 6570–6575 (2004).
8. Mootha, V. K. *et al. Nature Genet.* **34**, 267–273 (2003).
9. Lowell, B. B. & Shulman, G. I. *Science* **307**, 384–387 (2005).
10. Morino, K. *et al. J. Clin. Invest.* **115**, 3587–3593 (2005).

BUSINESS

Too little, too late?

Merger fever has finally reached Germany's drug firms. But can it cure the industry's woes? **Alison Abbott** reports.

Despite boasting an impressive historical heritage, Germany's once powerful pharmaceutical industry is now heavily fragmented. But the company that produced the world's first blockbuster drug is making a bid to reclaim some lost ground.

On 23 March, Bayer made an offer to buy Berlin-based Schering, quashing a hostile takeover bid from German rival Merck KGaA. Schering's board immediately announced its support for Bayer's €16.3-billion (US\$19.6-billion) offer. Darmstadt-based Merck, which is not linked to the New Jersey firm of the same name, withdrew its own offer of €14.6 billion for Schering the following day.

Bayer's chairman, Werner Wenning, describes the proposed merger as "the best way of reasserting the importance of Germany as a pharmaceutical industry base". But analysts are less convinced that the move will solve the problems faced by the companies.

The flurry of activity in Germany reflects the pressing need for mid-sized drug firms to attain critical mass, say analysts. None of the three companies involved is large enough to succeed in the current business climate, says Alexander Groschke, a pharmaceuticals analyst at Landesbank Rheinland-Pfalz in Mainz.

If Schering's shareholders accept the bid, as expected, then Bayer says it plans to axe 6,000 of the companies' combined workforce of just under 60,000. About a third of the losses are

likely to be in research and development, it says.

Günter Stock, president of the Berlin Brandenburg Academy of Sciences and a former head of research at Schering, criticizes the planned job cuts. "There have to be some financial savings accompanying the merger," he says. "But I cannot imagine that there will be as many as 6,000 jobs lost. Big cuts in the core business of Schering would endanger the future of Bayer."

Germany's drug sector was mostly bypassed by the merging frenzy of the 1990s that created the likes of AstraZeneca and GlaxoSmith-Kline. Today, the country has to compete with the resultant global entities. Bayer says that the merger will "create a healthcare heavyweight of international standing". But analysts say that mid-sized companies such as Schering, Merck and Bayer face immense challenges in keeping up with the increasingly global pharmaceuticals business.

This month, for example, consultants Wood Mackenzie of Edinburgh published a report warning such companies to act swiftly to protect themselves. As pipelines of potential drugs dry up, the pressure to license new products from small biotechnology companies is increasing, the report notes, which is pushing the price of such deals to levels that only the larger firms can afford. In the past five years, Wood Mackenzie says, the typical cost of such a deal has increased sevenfold.

Mid-sized companies have only three ways to respond, says the report: they can increase in size, specialize geographically or concentrate in a particular therapeutic area.

Schering is a world leader in contraceptives and other women's health products, and has enjoyed considerable success with Betaferon, its drug for multiple sclerosis. But last year alone it lost four major drug candidates during the final phase of clinical trials. It recently shut down its cardiovascular research and restricted its brain

research programme to multiple sclerosis and Parkinson's disease. And it has said that, apart from hormone therapies, the main focus of its research will be finding cancer therapies.

Leverkusen-based Bayer is probably best known as the company that produced aspirin, the first blockbuster drug. It has more diverse interests than Schering — only half of its activities are in pharmaceuticals; one-third is agricultural products and the rest is materials. Although it is strong in over-the-counter therapies, Bayer's healthcare business took a major hit in 2001 when its cholesterol-lowering drug Lipobay (or Baycol) was withdrawn after it was linked to patient deaths. Nevertheless, the company has had some recent success with cancer treatments, and last December US regulators approved its drug Nexavar (sorafenib) for treating kidney cancer.

If the merger goes ahead, the new company, Bayer-Schering Pharmaceuticals, will be based in Berlin and will rank as the world's 12th-largest pharmaceutical company. But Groschke argues that the price Bayer is paying for Schering is on the high side, and may reflect the company's fear of being left behind in a consolidation rush.

"With a €9-billion turnover, it would be at the high end of what is viewed as a mid-sized company," says Patrik Frei, chief executive of Venture Valuation in Zurich. "It would have more purchasing power to compete for biotech licensing deals." Shares in Bayer rose modestly on news of the merger.

Bayer-Schering Pharmaceuticals would have a reasonably strong focus on cancer. Across the board, it would have 50 or so drugs in the pipeline — with 19 in the last phases of clinical testing.

The firm would also have a strong geographical focus: Europe accounts for half the sales of both Bayer and Schering. This could be a mixed blessing, however, as Europe's price controls and diffuse healthcare systems make it a tough market for drug companies.

Frei says that the new company would have to invest smartly with biotech partners — perhaps even before a product enters clinical trials — in order to prosper. "It's a hard time for mid-sized pharmaceutical companies," he says. ■

M. SCHREIBER/AP



Schering is braced for job losses following Bayer's takeover bid.

Shared data are key to beating threat from flu

SIR — We fully support Ilaria Capua in her call for avian-influenza researchers to release data to the public, rather than store them in restricted databases, as reported in your Editorial “Dreams of flu data” (*Nature* **440**, 255–256; 2006). Keeping sequences secret, whatever the motivation, slows down scientific progress and hinders efforts to protect public health. The influenza genome sequencing project (www.niaid.nih.gov/dmid/genomes/mscs/influenza.htm) has, in the past year, sequenced more than 1,000 complete genomes of human influenza and released them to GenBank (www.ncbi.nih.gov/Genbank). All sequences are deposited immediately they are completed, as agreed by all the centres contributing samples to this project. We believe unrestricted access to these data will jump-start research in many influenza labs across the globe, advancing vaccine design and enhancing our understanding of the virus.

We call on all other scientists who might be sitting on influenza-virus data, whether human or animal data, to follow this example. We also join Capua and *Nature* in calling for the World Health Organization and the US Centers for Disease Control and Prevention to make future — and archived — data available to the scientific community. It is time for the community of influenza researchers to recognize, as the human genome sequencing project did ten years ago, that immediate public release of sequence data provides the greatest benefits to human health. The influenza virus does not respect national or other artificial boundaries, and we all need to work together to control it.

Steven Salzberg*, **Elodie Ghedin†**, **David Spiro‡**

*Center for Bioinformatics and Computational Biology, University of Maryland, College Park, Maryland 20742, USA

†The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, Maryland 20850, USA

Dogs can play useful role as sentinel hosts for disease

SIR — News that Thai dogs have tested positive for antibodies to the influenza A H5N1 virus (“Thai dogs carry bird-flu virus, but will they spread it?” *Nature* **439**, 773; 2006) reinforces our notion that carnivore and scavenger species have the potential to act as important sentinel hosts for emerging human and livestock diseases, providing a valuable tool for surveillance and for determining spatial and temporal patterns of infection.

Domestic dogs may prove particularly

useful as sentinel hosts, especially in developing countries. They are ubiquitous, with one dog for every 7 to 21 people in most parts of Africa and Asia. Dogs are known to be susceptible to a wide range of emerging human infections, and, as free-roaming scavengers in many parts of the world, they effectively ‘sample’ widely from a community environment. Despite appearances, domestic dogs in most developing countries are generally accessible for safe handling and sampling.

Our experience in Africa and Asia suggests that sampling dogs for disease surveillance would be particularly cost-effective if carried out in combination with rabies vaccination campaigns, as this provides owners with a strong incentive to participate. During these campaigns, several hundred dogs per day could be accessible for sampling at a cost of US\$1–2 per dog vaccinated (K. Bögel and F. X. Meslin *Bull. World Health Organ.* **68**, 281–291; 1990).

Domestic dogs, like other carnivore and scavenger species, may act as ‘bioaccumulators’ of pathogen exposure, with consumption of infected host material resulting in high rates of seroconversion. We suggest that they could therefore usefully be included as part of surveillance strategies to increase the efficiency of pathogen detection, particularly for pathogens that occur at low prevalences in animal reservoirs or are maintained in wild animal populations that are difficult to sample.

Age-seroprevalence data can also allow timing of outbreaks to be established retrospectively and with reasonable accuracy, for at least a number of years. This would be particularly valuable in areas where reporting and laboratory confirmation of human and animal disease outbreaks are limited, which may apply in many parts of the developing world.

Sarah Cleaveland*, **Francois X. Meslin†**, **Robert Breiman‡**

*Wildlife and Emerging Disease Section, Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush, Roslin, Midlothian EH25 9RG, UK

†Department of Food Safety, Zoonoses and Foodborne Diseases, World Health Organization, 1211 Geneva 27, Switzerland

‡International Emerging Infections Program, Centers for Disease Control and Prevention, PO Box 606, Village Market, Gigiri, Nairobi, Kenya

Ecological society supports its African counterparts

SIR — Your Editorial “It’s academic” (*Nature* **439**, 762–764; 2006) reports the need for stronger national academies in Africa. The British Ecological Society launched its Building Capacity for Ecology Fund in

January 2006, to support the development of ecological societies in Africa and Eastern Europe. The society has put £500,000 (US\$70,000) of its own money into this exciting initiative. As president (J.H.L.) and past-president (A.H.F.) of the society, we hope that others will join us in this or similar initiatives, as it is vital that the science of ecology is promoted and used by policy-makers in Africa.

John H. Lawton, **Alastair H. Fitter**

British Ecological Society, 26 Blades Court, Deodar Road, London SW15 2NU, UK

Populations who test drugs should benefit from them

SIR — Paul Herrling, in his Commentary article “Experiments in social responsibility” (*Nature* **439**, 267–268; 2006), describes pharmaceutical companies moving towards a more progressive approach to drug development and distribution in poor countries. But it is important to note that, even when research in developing nations leads to effective treatments, there is still the danger of local populations being exploited.

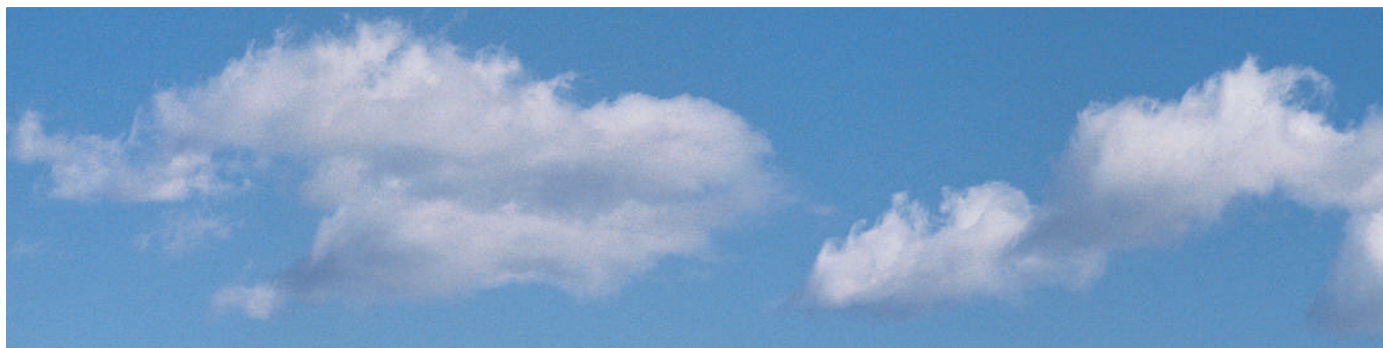
Recent clinical trials of a hepatitis E vaccine in Nepal are a case in point. Run by Glaxo-SmithKline and the US government Armed Forces Research Institute of Medical Sciences, the trials showed an impressive on-treatment efficacy for the experimental vaccine (see M. P. Shrestha and R. N. Scott’s report to the American Society for Tropical Medicine and Hygiene at www.astmh.org/meetings_new/ASTMH_05_FP2.pdf). But the methodology of the trial raises questions about the ethics of clinical-trial conduct among vulnerable populations. The research team had to drop original plans to test civilian volunteers in the city of Lalitpur, after local people objected to a lack of informed consent or participation in trial design (for details of these events, see J. Andrews *Am. J. Bioethics* **5**, W1; 2005). Instead, they gave the experimental vaccine to soldiers in the Royal Nepalese Army, who are vulnerable as members of the armed forces and as some of the poorest people in a ‘least-developed’ country.

These ethical issues take on greater importance now that the hepatitis E vaccine may have public-health usefulness. Will the Nepalese community benefit? Or will the results be used only to develop a profitable vaccine for (mainly Western) travellers and US soldiers? We hope that GlaxoSmithKline and its collaborators make this vaccine accessible to the populations placed at risk by the trial, in line with the well-intentioned humanitarianism that Herrling describes.

Sanjay Basu, **Jason Andrews**, **Duncan Smith-Rohrberg**

Yale University School of Medicine, 129 York Street, New Haven, Connecticut 06511, USA

BOOKS & ARTS



Blue-sky research

How scientists sought to explain the colour of the heavens.

Sky in a Bottle

by Peter Pesic

MIT Press: 2005. 256 pp. \$24.95, £16.95

Richard P. Wayne

The central question that *Sky in a Bottle* by Peter Pesic seeks to answer is why the cloudless daytime sky is blue. The bottle in the title alludes to experiments that have been performed over the ages in an attempt to capture or explain the blue colour. This all sounded tempting to me, as the past 45 years of my professional life have been devoted to a parallel enquiry, that of atmospheric chemistry as studied 'in a bottle'.

It is humbling to find in the book a diagram of a photochemical flow apparatus that John Tyndall used to produce and study photochemical smog in 1871; most present-day atmospheric chemists doubtless believe that such experiments are a modern development. One particularly nice touch is the inclusion of recipes for some of the most important 'sky in a bottle' experiments, suitably modified by Pesic to conform to present-day ideas about safety. Several of the earlier experiments used liquids rather than gases, but some provided clear hints that selective scattering of radiation, rather than selective absorption, might provide the key to understanding the sky's colour.

Pesic, a musician who holds a doctorate in physics, sets out on an enthralling and entertaining journey. He follows in the footsteps of the many artists, philosophers and mystics who have taken an interest in the blue of the sky, thus complementing the curiosity of scientists. Artists who appear in the book range from Leonardo da Vinci to Wassily Kandinsky, whose love of the colour brought the Blue Rider group of artists into being. Curiously, the sky does not seem to have been painted blue much before the time of Giotto's fresco

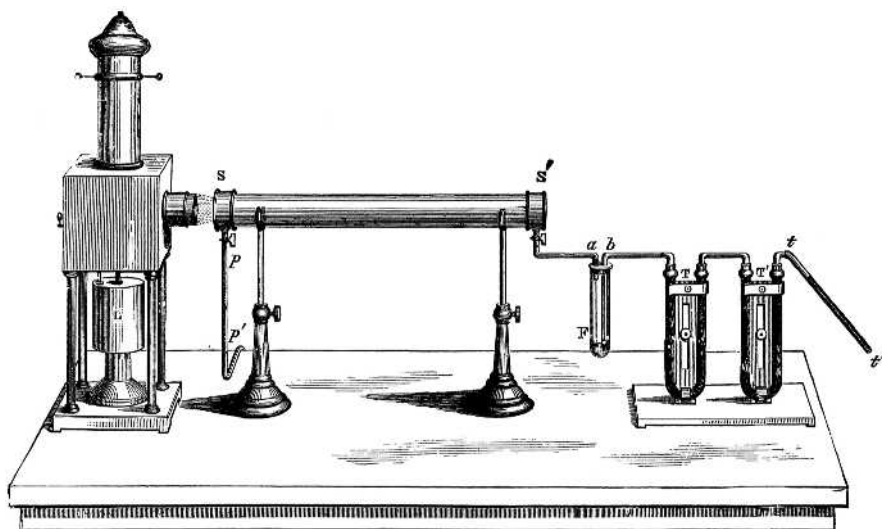
in Padua in 1305–6. This late arrival of blue was partly, but not entirely, the result of a lack of affordable blue pigments; ground-up lapis lazuli, which Giotto used, was ruinously expensive and came from 'beyond the seas' (hence 'ultramarine').

Starting in the tenth century, long after Pesic begins his tale, some of the scientists who made notable achievements include the founder of modern optics, Ibn al-Haytham, Roger Bacon, Leonardo again, René Descartes, Isaac Newton and Robert Hooke (who sometimes seemed to see more clearly than Newton). Thomas Young contributed with his use of slits, and William Herschel (originally a musician) and his son John played their part, as did Augustin Fresnel, Siméon-Denis Poisson and Jean Perrin. More recent additions to the list are James Clerk Maxwell, Lord Kelvin, the Rayleighs, Einstein and Max Planck.

Bit by bit, it became clear that the sky's blue

colour resulted from some kind of phenomenon involving the scattering of light. But what could be doing the scattering? Particles, presumably, but of what? Dust or water were initially favoured. In this context, it is noteworthy how many of the artists, philosophers and scientists were enthusiastic mountaineers or lovers of mountains, at a time when few people thought mountains had any beauty or interest. They, of all people, must have known that the sky seems more intensely blue when the atmosphere is dry and dust free. Even more curiously, evidence accumulated that any scattering particles must have the same refractive index as air itself.

Since the time of Newton, the wave theory of light has gradually become widely accepted, thanks to its power to explain interference phenomena; there has been a growing understanding of polarization; and doubters have been converted to the reality of atoms and



John Tyndall passed light through filtered air and vapour from a volatile liquid to create 'sky in a bottle'.

molecules. The cornerstone of today's view, developed by the third Baron Rayleigh, is that light is scattered by the molecules of the air, principally nitrogen and oxygen. The interaction of these molecules with electromagnetic radiation leads to some of the light being diverted from its original path and re-radiated isotropically. This process is known as Rayleigh scattering, and its efficiency increases as the fourth power of the frequency — very sharply. Wavelength is inversely proportional to frequency, so shorter wavelengths are scattered much more than longer ones: blue much more than red.

Marian Smoluchowski reached similar conclusions in a different way, basing his argument on the opalescence of carbon dioxide near its critical point, as observed by Einstein and Thomas Andrews. One last point worth making is that the scattering in our atmosphere that makes the sky blue requires the molecules

to act independently of each other — the scattering has to be incoherent. If the density of the atmosphere were too high, the scattering would be multiple and coherent (the latter occurs in transparent solids), so the colour would be lost. With too low a density, there would be too little scattering to be registered by the eye.

Pesic has a broadly based erudition and the science in his book is pretty sound. I do not agree entirely with his account of the excitation of the airglow, and I know that Earth is not 10 billion years old, but these are small matters. A huge amount of important science is presented in an accessible way that makes no huge demands on non-scientist readers, and the level and approach provide enough intellectual substance for the specialist. Copious notes flesh out many of the points made in the main text. The notes are preceded by an interesting selection of letters between George

Stokes, Tyndall and Lord Kelvin on the blue of the sky. Many of the concepts developed are taken for granted by almost everyone and are deeply ingrained in the very heart of modern science. Science does not always proceed in a linear or expected fashion, being helped forward in some cases by inspiration. As Pesic says of a prose poem by Edgar Allan Poe dedicated to Alexander von Humboldt: "a visionary was able to discern a conclusion vindicated only much later by sober science".

I commend this book to those who want to read about truly significant discoveries linked together through a need to answer what seems to be a simple question. Unlike many other attempts to popularize science, this book has managed not to garble the facts or sensationalize them. It is well worth reading. ■

Richard P. Wayne is in the Department of Chemistry, University of Oxford, South Parks Road, Oxford OX1 3QZ, UK.



Smooth operator: a man takes advantage of the properties of ice to hitch a lift in 1930s Paris.

The big freeze

Ice: The Nature, the History and the Uses of an Astonishing Substance

by Mariana Gosnell

Alfred A. Knopf: 2005. 576 pp. \$30

Greg Dash

In its various forms, solid H₂O covers large parts of Earth's surface for all or much of the year, with implications for our climate. The reflectivity of snow and sea ice greatly influences the planet's heat budget, with any reduction in coverage providing a positive-feedback mechanism for global warming. In her book *Ice*, Mariana Gosnell quotes some researchers' gloomy predictions: "Nobody knows how to stop the warming of the Earth." Yet there are enough interactions in the system for some effects, such as increased precipitation, to perhaps counteract the warming. Still, Gosnell's

book isn't only about the serious topic of global warming; instead she takes a much lighter look at the solid phase of water.

Ice has such diverse properties that it cuts across a range of disciplines. Gosnell's 36 chapters describe the basic science, the importance and the drama for specialists of each, be they glaciologists, atmospheric scientists or students of permafrost, ice coring and ice physics.

One of the most unusual applications was a project to build an unsinkable aircraft carrier. During the early days of the Second World War, Nazi submarines sank a lot of ships crossing the Atlantic from the United States to Britain. Geoffrey Pyke, who was director of programmes for Lord Mountbatten, proposed that a great artificial iceberg be built to serve as a floating base in the Atlantic for submarine-hunting planes. One of the essential

features of the project was that the ice would be strengthened with wood pulp. In Pyke's honour, the composite was christened 'pykrete'. Churchill liked the idea, and the project began, with a test model built in Canada. But the project never progressed very far, being made redundant by Allied aircraft with increased range. Max Perutz, who later received a Nobel prize for his work in biochemistry, was a young researcher on the project. Years later he wrote: "I think that had not the course of the war and the state of our armaments changed, the bergship could have been constructed." It would surely have been the world's largest ice sculpture.

Gosnell writes of the start of her interest in ice. She had been sent as a journalist to report on an expedition to the Canadian high Arctic, where tests were being done to quantify the damage that floating ice might cause to tankers or oil rigs. Intrigued by what she learned, she began to read a lot about ice. The book's bibliography is impressive, and I get the impression she has read a good deal of the listed articles. Most of the chapters are written with clear descriptions and understanding.

At least one of the chapters comes out of Gosnell's personal experience. She writes: "One winter I rented a cabin in Elkins, New Hampshire, to watch a lake freeze." After waiting through windy days and fluctuating temperatures, the event finally happens. "In one small, secluded spot... I see, in a bay where the beach slopes gently into the lake and the water is only a few inches deep, what looks like a windowpane. I am excited out of all proportion to the event. First lake ice!"

All the chapters are written with similar life and imagery. As a result, the book will make for enjoyable reading by specialists and non-specialists alike. ■

Greg Dash is emeritus professor in the Department of Physics, University of Washington, Seattle, Washington 98195, USA.

A selfish element

Genes in Conflict: The Biology of Selfish Genetic Elements

by Austin Burt & Robert Trivers

Belknap Press: 2006. 610 pp. \$35, £21.95

James F. Crow

In the early 1970s, Robert Trivers published six articles on the evolution of social behaviour. They were ignored by most social scientists, who were reluctant to consider natural selection as a cause of human behavioural traits, and they were bitterly attacked by marxists for reasons of doctrine. Some population geneticists also objected, because Trivers did not formulate his concepts as standard equations for changes in allele frequency. Yet the papers have proved seminal, and Trivers is now recognized, along with Edward O. Wilson and Bill Hamilton, as one of the founders of behavioural evolution and sociobiology. A collection of his papers has been published (*Natural Selection and Social Theory*; Oxford University Press, 2002), along with personal histories that add greatly to their interest. Young scientists will be empathetic: both Hamilton and Trivers had trouble with publishers, grants and critics.

Just over ten years ago, Trivers joined forces with geneticist Austin Burt for a detailed study of selfish genetic elements, and *Genes in Conflict* is the result of their fruitful collaboration. The book is the first of its kind and admirably fills an empty niche.

The mechanism of meiosis strongly selects against elements that violate Mendel's rules, yet there are clever tricksters that manage to beat the system. Some of these are well known, but it is surprising how many there are. There seems to be no molecular or cellular process that cannot be subverted by a sufficiently clever opportunist. The book starts with the well-studied *t*-alleles in the mouse and segregation distortion in fruitflies. Successive chapters then treat selfish sex chromosomes, imprinting, selfish mitochondria, gene conversion, 'homing', transposable elements, B chromosomes, and the exclusion of whole genomes. The book includes some 1,500 references, and several of the scientists who have contributed most to evolutionary explanations are mentioned, of whom David Haig stands out for his inventiveness.

Each of these subjects is treated fully, with their detailed cellular mechanisms described and, when they are known, the molecular explanations. An important feature of the book is that each subject is treated from an evolutionary viewpoint, as you might expect from these authors. Their views are imaginative but are sometimes speculative or difficult to test; they are great if you don't inhale. Yet for me, this is what makes the book especially interesting — it is far more stimulating than a

series of descriptions. Tables that summarize the facts will help readers keep the details straight, and the many diagrams will aid their understanding of the mechanisms. Even so, I needed to have a pencil and paper handy. The authors also raise practical possibilities of the work, such as engineering driven Y chromosomes to control insect numbers.

The authors emphasize that the selfish entities they describe are almost always harmful to the host. They have little patience with those who assume an adaptive significance, including even Barbara McClintock, dismissing her idea that transposons — segments of DNA that can move around the genome — act as 'controlling elements'. Despite this, some elements have been domesticated and do, in fact, serve a useful purpose. Some of the most interesting speculation surrounds the possible origin of normal processes from selfish elements in the past. Two of the most convincing are mating-type switching in yeast and telomeres in several species.

The chapter on selfish cell lineages focuses on cancers as the most obvious example. The authors report the work of Andrew Wilkie and

colleagues, who explained the high mutation rate in Apert's syndrome (a condition involving the premature fusion of bones). It is caused by a switch in the male germ line from asymmetric stem-cell divisions, in which the daughter cells are different, to symmetric exponential growth, which greatly increases the number of mutant cells. This highlights the susceptibility of a cell line in which one daughter product undergoes apoptosis to subversion by exponentially growing variants. There must be an evolutionary trade-off between the advantages of asymmetric growth (providing a constant number of sperm continuously over a long lifetime) and a consequent high rate of deleterious mutation. Asymmetric cell division seems to be a prime target for subversion by exponentially growing derivatives. There are likely to be more examples waiting to be found, especially ones that do little harm.

The book ends with a thoughtful chapter on the present position and on future directions of study. Although many of the selfish elements so fully described in the book have lost their motility and died (sometimes increasing the size of the genome), the subject is very much alive.

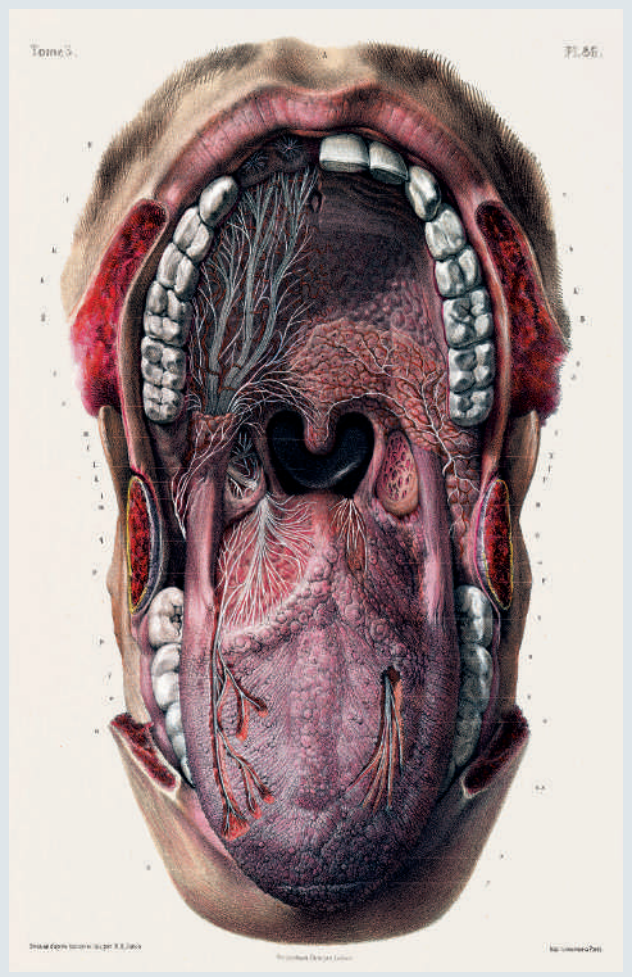
James F. Crow is professor emeritus of genetics at the University of Wisconsin-Madison, 425G Henry Mall, Madison, Wisconsin 53706, USA.

Inside knowledge

The science of anatomy reached a peak in nineteenth-century Paris, where anatomist Jean Baptiste Marc Bourgery and illustrator Nicolas Henri Jacob completed a monumental opus on anatomy and surgery.

The eight-volume work, which incorporates more than 700 illustrations of remarkable clarity and distinctive style, took two decades to complete.

The plates are now reproduced in their entirety in the *Atlas of Human Anatomy and Surgery* (Taschen, £100, \$200, €150), which has accompanying text in French, English and German. **A.A.**



TASCHEN

Quantum mechanics in the brain

Does the enormous computing power of neurons mean consciousness can be explained within a purely neurobiological framework, or is there scope for quantum computation in the brain?

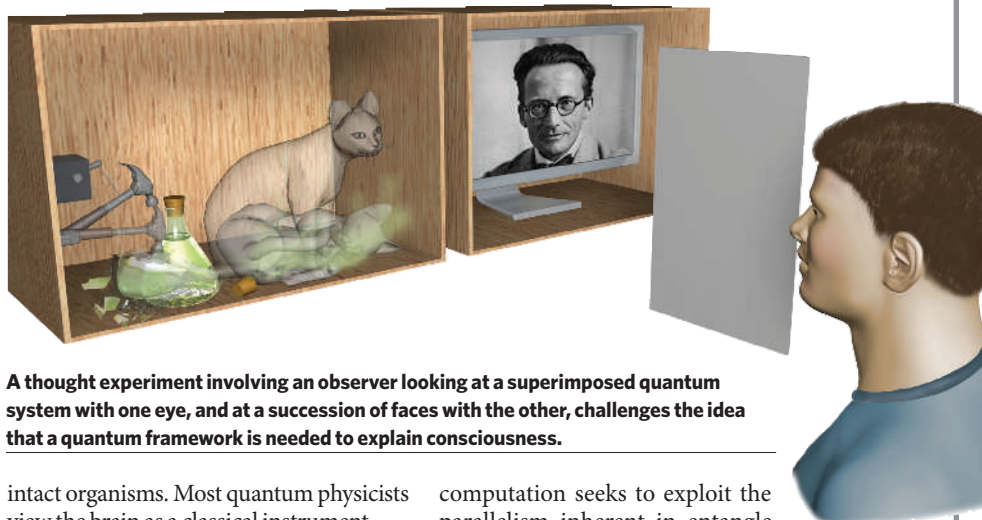
Christof Koch and Klaus Hepp

The relation between quantum mechanics and higher brain functions, including consciousness, is often discussed, but is far from being understood. Physicists, ignorant of modern neurobiology, are tempted to assume a formal or even dualistic view of the mind–brain problem. Meanwhile, cognitive neuroscientists and neurobiologists consider the quantum world to be irrelevant to their concerns and therefore do not attempt to understand its concepts. What can we confidently state about the current relationship between these two fields of scientific inquiry?

All biological organisms must obey the laws of physics, both classical and quantum. In contrast to classical physics, quantum mechanics is fundamentally indeterministic. It explains a range of phenomena that cannot be understood within a classical context: the fact that light or any small particle can behave like a wave or particle depending on the experimental setup (wave–particle duality); the inability to simultaneously determine, with perfect accuracy, both the position and momentum of an object (Heisenberg's uncertainty principle); and the fact that the quantum states of multiple objects, such as two coupled electrons, may be highly correlated even though the objects are spatially separated, thus violating our intuitions about locality (entanglement).

Major philosophical and conceptual problems surround the process of making measurements in quantum mechanics. To illuminate the paradoxical nature of superposition — that is, the fact that particles or quantum bits (qubits) are allowed to exist in a superposition of states — Schrödinger proposed a celebrated thought experiment: a sealed box containing the quantum superposition of both a dead and a live cat. When an observer peers inside the box, measuring its content, the wave function, which describes the probability that the system will be found in any one particular state, is said to collapse, and the system will be found in one or the other state with known probability.

The role of the conscious observer in this measuring process has been hotly debated since the early days of quantum mechanics. It is fair to say, however, that consciousness has been only a place holder in a chain of mathematical formulae, without much relevance to the study of neural circuits in



A thought experiment involving an observer looking at a superimposed quantum system with one eye, and at a succession of faces with the other, challenges the idea that a quantum framework is needed to explain consciousness.

intact organisms. Most quantum physicists view the brain as a classical instrument.

The critical question we are concerned with here is whether any components of the nervous system — a 300-degrees Kelvin tissue strongly coupled to its environment — display macroscopic quantum behaviours, such as quantum entanglement, that are key to the brain's function.

Specific molecular machines and proteins have been proposed to implement quantum computations. The best known of such proposals is Penrose and Hameroff's hypothesis that the tubulin components of microtubules, filamentous protein polymers that form the cytoskeleton of cells, implement quantum computations.

Lessons from quantum computers

When kings build, their followers team up in international journals and conferences, confusing the general public about the distinction between science and poetry. But large quantum systems are notoriously difficult to analyse rigorously, except in highly idealized models or limits. Estimates based on the same unrealistic one-particle model, applied to trillions of interacting particles, show discrepancies of ten orders of magnitude in the work of different authors. It is therefore better to turn to hard experimental realities and abstract computational theory to find the neural correlates of quantum processes in the brain.

Quantum computations are difficult to implement. In its simplest version, a quantum computer transforms the state of many two-dimensional qubits using a reversible, linear, probability-conserving mapping via a sequence of externally controllable quantum gates into a final state with a probabilistic outcome. Quantum

computation seeks to exploit the parallelism inherent in entanglement by assuring that the system is very likely to converge on the computationally desirable result.

This requires that the qubits are well isolated from the rest of the system. Coupling the system to the external world is necessary for the preparation of the initial state (the input); for the control of its evolution; and for the actual measurement (the output). However, all these operations introduce 'noise' into the computation (decoherence). Although some decoherence can be compensated for by redundancy and other fault-tolerant techniques, too much is fatal.

In spite of an intensive search by many laboratories, no scalable large quantum computers are known. The record for quantum computation is the factoring of the number 15 by liquid-state nuclear magnetic resonance (NMR) techniques. Qubits and a set of universal quantum gates have been proposed in many different implementations, but all solutions have serious drawbacks: photons interact only weakly with one another; nuclear spins in individual molecules are few in number in current devices, as are trapped atoms or ions. This paints a desolate picture for quantum computation inside the wet and warm brain.

Although brains obey quantum mechanics, they do not seem to exploit any of its special features. Molecular machines, such as the light-amplifying components of photoreceptors, pre- and post-synaptic receptors and the voltage- and ligand-gated channel proteins that span cellular membranes and underpin neuronal excitability, are so large that they can be treated as classical objects. (Their relative molecular masses range from 20,000 to 200,000; the two main

Q. PAUL/AIP

ESSAY

dimers of tubulin are around 55,000.)

Two key biophysical operations underlie information processing in the brain: chemical transmission across the synaptic cleft, and the generation of action potentials.

These both involve thousands of ions and neurotransmitter molecules, coupled by diffusion or by the membrane potential that extends across

tens of micrometres. Both processes will destroy any coherent quantum states. Thus, spiking neurons can only receive and send classical, rather than quantum, information. It follows that a neuron either spikes at a particular point in time or it does not, but is not in a superposition of spike and non-spike states.

The power of quantum mechanics is often invoked for problems that brains solve efficiently. Computational neuroscience is a young field and theories of complex neural systems, with all the variability of living matter, will never reach the precision of physical laws of well-isolated simple systems. It has already been demonstrated, however, that many previously mysterious aspects of perception and action are explainable in terms of conventional neuronal processing.

Two examples are models for the rapid recognition of objects (for example, animals or faces) in natural scenes, with performance approaching that of human observers, and the attentional selection of objects in cluttered images. The necessary mathematical operations — such as changes in synaptic weights, evaluating the inner product between presynaptic activity and synaptic weight, multiplication and stationary nonlinearities — are available to neurons. Indeed, there is an *embarras de richesse* of computational primitives implemented by synapses, dendrites and neurons. That is not to suggest that we understand how brains compute. But so far, there seems to be no need for quantum skyhooks.

The reason for the unprecedented computational power of nervous systems is their high degree of parallelism. For instance, filter-like operations in retinal or cortical cells in the visual stream are performed simultaneously on an entire image and thus are not limited by the tyranny of a single processor. Furthermore, unlike the von Neumann architecture of the programmable digital computer, the brain intermixes memory elements in the form of modifiable interconnections within the computational substrate, the neuronal membrane. Thus, no separate memory 'fetch' and 'store' cycles are necessary.

Much of the hope that quantum mechanics works in the brain is pinned to the supposition that quantum algorithms, which

are much more powerful than conventional algorithms (based on classical physics), are implemented in the nervous system. The most famous of these is Shor's procedure for factoring large integers for data encryption.

However, in the past decade no quantum algorithm of similar power and applicability to Shor's has been found. And factoring large numbers is not something for which the brain has much use.

Why should evolution have turned to quantum computation, so fickle and capricious, if classical neural-network computations are evidently entirely sufficient to deal with the problems encountered by nervous systems?

Food for thought

At this point, intrepid students of the mind point to qualia, the constitutive elements of consciousness. The subjective feelings associated with the redness of red or the painfulness of a toothache are two distinct qualia. As long as it remains mysterious how the physical world gives rise to such sensations, could one of the more flamboyant interpretations of quantum mechanics explain consciousness? Most provocatively, Roger Penrose has claimed that brains can evaluate noncomputable functions; that this ability is related to consciousness; that both this ability and consciousness require a yet-to-be-discovered theory of quantum gravity and that microtubules are the sites of the associated quantum gates.

The problem of consciousness and its neuronal correlates is beginning to emerge in outlines. The content of consciousness is rich and highly differentiated. It is associated with the firing activity of a very large number of neurons spread all over the cortex and associated satellites, such as the thalamus. Thus, any one conscious percept or thought must be expressed in a wide-flung coalition of neurons firing together. Even if quantum gates exist within the confines of neurons, it remains totally nebulous how information of relevance to the organism would get to these quantum gates. Moreover, how would it be kept coherent across the milli- and centimetres separating individual neurons when synaptic and spiking processes, the primary means of neuronal communication on the perceptual timescale, destroy quantum information?

It is far more likely that the material basis of consciousness can be understood within a purely neurobiological framework, without invoking any quantum-mechanical *deus ex machina*.

We challenge those who call upon consciousness to carry the burden of the measurement process in quantum mechanics

with the following thought experiment. Visual psychology has caught up with magicians and has devised numerous techniques for making things disappear. For instance, if one eye of a subject receives a stream of highly salient images, a constant image projected into the other eye is only seen infrequently. Such perceptual suppression can be exploited to study whether consciousness is strictly necessary to the collapse of the wave function.

Say an observer is looking at a superimposed quantum system, such as Schrödinger's box with the live and dead cat, with one eye while his other eye sees a succession of faces (see figure). Under the appropriate circumstances, the subject is only conscious of the rapidly changing faces, while the cat in the box remains invisible to him. What happens to the cat? The conventional prediction would be that as soon as the photons from this quantum system encounter a classical object, such as the retina of the observer, quantum superposition is lost and the cat is either dead or alive.

This is true no matter whether the observer consciously saw the cat in the box or not. If, however, consciousness is truly necessary to resolve the measurement problem, the animal's fate would remain undecided until that point in time when the cat in the box becomes perceptually dominant to the observer. This seems unlikely but could, at least in principle, be empirically verified.

The empirical demonstration of slowly decoherent and controllable quantum bits in neurons connected by electrical or chemical synapses, or the discovery of an efficient quantum algorithm for computations performed by the brain, would do much to bring these speculations from the 'far-out' to the mere 'very unlikely'. Until such progress has been made, there is little reason to appeal to quantum mechanics to explain higher brain functions, including consciousness. ■

Christof Koch is in the Division of Biology and the Division of Engineering and Applied Science, 216-76, California Institute of Technology, Pasadena, California 91125, USA.

Klaus Hepp and Christof Koch are at the Institute of Neuroinformatics, the University of Zürich and ETH, Zürich, Switzerland.

FURTHER READING

Hepp, K. in *Quantum Future: Lecture Notes in Physics* (eds Blanchard, P. & Jadczyk, A.) **517**, 92-104 (1998).
Koch, C. *Biophysics of Computation: Information Processing in Single Neurons* (Oxford Univ. Press, New York, 1999).
Koch, C. *The Quest for Consciousness: A Neurobiological Approach* (Roberts, Colorado, 2004).
Nielsen, M. & Chuang, I. *Quantum Computation and Quantum Information* (Cambridge Univ. Press, Cambridge, 2002).
Penrose, R. *The Emperor's New Mind* (Oxford Univ. Press, Oxford 1989).

NEWS & VIEWS

ECOLOGY

Green and pleasant trials

Peter D. Moore

In the 1980s, a large lake — Lago Guri — was created as part of a hydroelectric project in Venezuela. Islands in the lake have enabled ecologists to test a fundamental hypothesis in their discipline.

Why is the world green? Why have grazing animals with their insatiable appetites not consumed all vegetation and reduced the land to dust? There have been hypotheses, of course, but as with many large-scale ecological problems, it has not proved easy to test any proposal with controlled experiments. One suggestion is that the intensity of grazing is held in check by predation of carnivores on the herbivores, and this hypothesis has at last proved testable. Writing in *Journal of Ecology*, John Terborgh and his colleagues¹ describe a large-scale experiment in which the degree of predation upon grazers varies and the consequences for vegetation can be measured. They show that, without top predators, the world would be less likely to remain a green and pleasant land.

Animal life is supported by the primary production of green plants, and current knowledge² suggests that for every species of terrestrial plant there are about five species of animal. Undoubtedly, many more species of animal (especially insects) await description than do plants. Not all of these animals feed directly

on living plants; some consume dead plant litter, and others prey on the plant consumers. But in the light of such energetic dependency of animals on a plant food-base, it is remarkable that vegetation survives at all — and not only survives, but dominates the biomass of most land ecosystems. The most widely accepted explanation for this, first put forward by Hairston *et al.*³, is that herbivore numbers are controlled by ranks of predators that keep their populations in check and inadvertently ensure that green plant production continues.

An opportunity to test the hypothesis on a meaningful scale arose when a valley in Venezuela was flooded to develop a hydroelectric scheme, and a lake — Lago Guri (Fig. 1) — was created. The lake is 4,300 km² in area, and contains many islands of different sizes. Before the valley was flooded, commercial logging of the valley floor was carried out, but the elevated regions were left untouched and survive as forested islands. Terborgh *et al.* have recorded the ecological consequences of fragmentation of the forest into these isolated

units over many years⁴, and have described the relationship between island size and species richness, which follows the model described by the theory of island biogeography⁵. Species losses, predictably, have been greater on the small islands.

Islands of less than 2 hectares (20,000 m², or about 5 acres) lost many of their vertebrate species within a few years of isolation, and these smaller islands also began to display higher densities of herbivores⁶ — especially invertebrates, including leaf-cutter ants, but also some vertebrates such as iguana, howler monkey, agouti and tortoise. Land masses of more than 75 ha retained greater numbers of vertebrate grazers, including deer, peccary and a full range of primates, but they also supported predators of these vertebrates, including raptors (such as harpy eagle), snakes, ocelot, puma and jaguar. The Hairston 'green world' hypothesis would predict that the very small islands that lacked predators and developed high densities of herbivores should experience a decline in vegetation. Medium-sized islands (less than 15 ha) with some vertebrate

Figure 1 | The islands of Lago Guri.



PETER LANGER ASSOCIATED MEDIA GROUP

predators, such as the armadillo that preys on leaf-cutter ants, would be less severely affected, and large islands with a full complement of predators would remain unchanged.

Terborgh's team periodically surveyed the vegetation of all three types of island. The small islands typically contained about 300 individual trees, so all of these were tagged and their sizes and condition noted. Sample areas (usually about 0.6 ha in extent) with similar tree densities were selected on the medium-sized and larger islands, and the individual trees were recorded in the same way. Changes rapidly became evident on the small islands, which by 1997 had densities of small saplings only 37% of those on the large islands; recruitment and mortality of trees and shrubs had evidently been strongly affected by the increased herbivory under conditions of low predation. By 2002, the density figure for small islands had fallen to 25% of that of the large islands. Tree and shrub mortality over a five-year period was quite high on all islands, but was greatest on the small ones, which experienced 46% mortality compared with 32% on the large islands.

The researchers consider other causes, but conclude that the loss of animals that preyed upon vertebrate grazers and leaf-cutter ants on the small islands set in motion a trophic cascade that destabilized the food web. Such cascades, where the removal of one trophic level (in this case, top predators) causes knock-on effects through other trophic levels, are well documented from aquatic communities⁷. They have proved difficult to demonstrate in terrestrial ecosystems, although (for example) the loss of wolves from most of the national parks of the United States has led to increases in vertebrate grazers and overgrazing.

Terborgh *et al.*¹, however, have quantified these effects with great precision and have demonstrated both the extent and pace of the trophic cascade. It remains to be seen whether overgrazing will lead to the total destruction of vegetation on the small islands, and whether that would then lead to herbivore extinction followed by plant reinvasion and the establishment of a new order.

Peter D. Moore is in the Department of Biochemistry, King's College London, Franklin Wilkins Building, 150 Stamford Street, London SE1 9N, UK.
e-mail: peter.moore@kcl.ac.uk

1. Terborgh, J., Feeley, K., Silman, M., Nuñez, P. & Balukjian, B. *J. Ecol.* **94**, 253–263 (2006).
2. Groombridge, B. (ed.) *Global Biodiversity: Status of the Earth's Living Resources* (Chapman & Hall, London, 1992).
3. Hairston, N. G., Smith, F. E. & Slobodkin, L. B. *Am. Nat.* **94**, 421–424 (1960).
4. Terborgh, J. *et al. Science* **294**, 1923–1926 (2001).
5. MacArthur, R. H. & Wilson, E. O. *The Theory of Island Biogeography* (Princeton Univ. Press, 1967).
6. Rao, M., Terborgh, J. & Nuñez, P. *Conserv. Biol.* **15**, 624–633 (2001).
7. Paine, R. T. *J. Anim. Ecol.* **49**, 667–685 (1980).

PLANETARY SCIENCE

Saturn's bared mini-moons

Frank Spahn and Jürgen Schmidt

Propeller-shaped structures seem to reveal the presence of moonlets, about 100 metres in diameter, embedded in Saturn's rings. This discovery adds to our picture of how the rings formed and are evolving.

The question of where Saturn's magnificent system of rings came from has intrigued planetary scientists for centuries. A currently favoured thesis is that the flat disk of the main rings, which girdle the planet's equator, originated in the dispersion of material from the disruption of an icy satellite following the impact of a comet or asteroid^{1,2}. Such a giant impact would have left behind debris in a broad range of sizes. But apart from two moons of kilometre size, only a main population of ice particles from a few centimetres to a few metres across has so far been deduced from remote sensing³. The detection of propeller-shaped brightness undulations in the rings, reported by Tiscareno *et al.* on page 648 of this issue⁴, supplies the first evidence for large ring particles of between 40 and 120 metres in diameter. Their discovery bridges the size gap between the main population and the embedded moons.

The images on which Tiscareno and colleagues base their analysis were taken by the Cassini spacecraft, which is currently investigating the Saturn system. Two fundamental physical processes within Saturn's rings allow an embedded large boulder, or moonlet, to generate the kind of structure that the authors detect: gravity and collisions. Moonlet and ring particles both orbit in the strong gravity field of Saturn, so their mutual gravitational

attraction will, contrary to intuition, act to scatter the particles away from the moonlet. So gravity tends to clear a gap around the orbit of the moonlet, and the width of this gap is proportional to the moonlet's size.

This process is, however, counteracted by frequent collisions among ring particles — typically 10 to 100 per orbital revolution of the rings, lasting about 10 hours — that jostle particles from high-density regions to the gravitationally depleted gaps. The stationary pattern that emerges between these two processes will depend on the size of the moonlet and the number density of the ring particles. If a body embedded in Saturn's A ring (the outer of the planet's two brightest rings, A and B) is larger than about 1 kilometre in diameter, its gravity will be strong enough to keep open a directly detectable gap around the ring's entire circumference. But for smaller moonlets, diffusion of particles as a result of collisions will close the gap at some distance from the moonlet. An incomplete, asymmetric gap, flanked by density enhancements, forms (Fig. 1). This is the origin of the propeller pattern observed by Tiscareno *et al.*⁴ (Fig. 1 on page 649).

The propellers offer a unique chance to estimate the number of such embedded moonlets. Boulders 100 metres in diameter are too small to be seen directly, and because they are too rare to affect the optical appearance of the rings

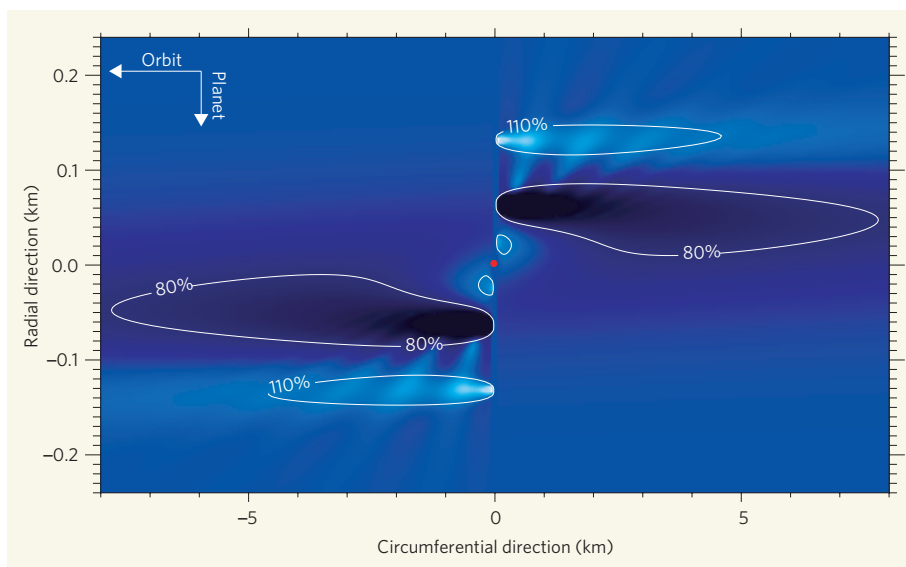


Figure 1 | Moonlet and propeller. The propeller structure induced in a model¹¹ by a 40-metre-diameter icy moonlet in Saturn's rings (marked by red dot). Dark colour corresponds to density depletion of material, bright colour to balancing enhancement. Tiscareno and colleagues⁴ observe such structures in Cassini images of Saturn's rings.



Figure 2 | Saturn's rings. Processes of accretion and fragmentation of ring particles are emphasized. The boulder in the foreground accretes smaller ring particles through an S-shaped structure very similar to the propellers. (Artist's impression by W. K. Hartmann.)

collectively, their number cannot be inferred by photometry — the study of objects' brightness. But photometry can be used to obtain an idea of the distribution of sizes of the main particle population in Saturn's rings (those with radii ranging from centimetres to a few metres). The number of particles N with a radius greater than r is found to follow approximately an inverse-square law³, $N(>r) \sim r^{-2}$. This means that for each boulder with a diameter between 5 and 15 metres, there are about 100 particles of sizes between 0.5 and 1.5 metres, and 10,000 particles between 5 and 15 centimetres.

Looking at the number of gaps in the ring system, the number of kilometre-sized objects can also be inferred. There are two known moons embedded in the rings that plough circumferential gaps through the A ring: Pan (with a diameter of around 10 kilometres) in the 325-kilometre-wide Encke gap⁵ and Daphnis (diameter around 5 kilometres) in the 42-kilometre Keeler gap. Even though diffuse ringlets within the Encke gap⁶, and clear narrow gaps in the 4,800-kilometre Cassini division between the A and B rings⁷, imply the presence of further kilometre-sized moonlets, their number would be too small by far to be consistent with an extension of the inverse-square law for the sizes of the main population to the kilometre scale.

Interpolating between the number of 10-metre particles from photometric observations to the number of known kilometre-sized moons (two) would imply a size distribution in this region that falls off very steeply⁸, approximately as $N(>r) \sim r^{-4}$. (That exponent would mean that, for each moonlet in the size range between 0.5 and 1.5 kilometres, there are about 10,000 bodies with diameters between 50 and 150 metres, and 100 million between 5 and 15 metres.) Tiscareno and colleagues' observations⁴ are, taking into account the

statistical uncertainties, consistent with such a steep distribution (Fig. 3 on page 650).

The ring system's global distribution of particle sizes — including the embedded moons, the population of intermediate-sized boulders identified by Tiscareno *et al.*, and the main population of ring particles — provides evidence for processes of particle fragmentation and reaccretion in the rings that are probably still going on (Fig. 2). Following formation in the break-up of an ice moon, the primordial size distribution of the rings may have evolved to its present form by dint of such processes. Spectra of the rings at ultraviolet wavelengths⁹

also indicate relatively fresh water-ice in certain ring regions, implying that parts of the system are younger, perhaps recreated episodically by more recent moonlet disruptions.

The images in which the propeller structures were identified were taken from the unlit side of the rings as Cassini inserted itself into orbit around Saturn. Given the viewing geometry and illumination at the time, the high contrast of the propellers in these images is difficult to square with our current understanding. Photometric modelling of dynamic simulations¹⁰ might help here to define the particle properties better. The ring images from orbit insertion had the highest possible resolution in Cassini's nominal tour of the Saturn system. However, the higher inclinations of the spacecraft scheduled for late 2006 could provide favourable conditions for a systematic survey of larger propellers induced by the much less common moonlets that exceed a few hundred metres in size. Saturn's rings, long mysterious and compelling, may yet hold more secrets.

Frank Spahn and Jürgen Schmidt are in the Arbeitsgruppe Nichtlineare Dynamik, Universität Potsdam, Postfach 601553, 14469 Potsdam, Germany.

e-mail: frank@agnld.uni-potsdam.de

1. Dones, L. *Icarus* **92**, 194–203 (1991).
2. Colwell, J. E. *Planet. Space Sci.* **42**, 1139–1149 (1994).
3. French, R. G. & Nicholson, P. D. *Icarus* **145**, 502–523 (2000).
4. Tiscareno, M. S. *et al. Nature* **440**, 648–650 (2006).
5. Showalter, M. R. *Nature* **351**, 709–713 (1991).
6. Porco, C. C. *et al. Science* **307**, 1226–1236 (2005).
7. Lissauer, J. J. *et al. Nature* **292**, 707–711 (1981).
8. Cuzzi, J. N. *et al. in Planetary Rings* (eds Brahic, A. & Greenberg, R.) 73–199 (Univ. Arizona Press, Tucson, 1984).
9. Esposito, L. V. *et al. Science* **307**, 1251–1255 (2005).
10. Salo, H. & Karjalainen, R. *Icarus* **164**, 428–460 (2003).
11. Sremčević, M. *et al. Mon. Not. R. Astron. Soc.* **337**, 1139–1152 (2002).

NEUROSCIENCE

Rewinding the memory record

Laura L. Colgin and Edvard I. Moser

How does the brain store sequences of experience? Clues come from brain recordings of rats running along a track. The animals' memories seem to be consolidated in an unexpected way as they rest between runs.

Memories develop in several stages. After the initial encoding of new information during learning, memories are consolidated 'off-line', seemingly while not being actively thought about, through a cascade of events that is not well understood. In humans and other mammals, such an enhancement of recent memories may occur during sleep¹. But on page 680 of this issue, Foster and Wilson² show that substantial consolidation might also happen while awake during rest periods.

Insight into how sleep benefits memory consolidation has been gained by recording

neural activity in the hippocampus, a brain region that is crucial for mnemonic processing³. Cells that are activated in the hippocampus during certain awake behaviours fire in the same order but faster during the subsequent slow-wave phase of sleep^{4,5}. This reactivation of firing patterns occurs during 'sharp waves', excitatory waveforms that dominate hippocampal recordings throughout slow-wave sleep⁶. Sharp waves are accompanied by very fast oscillations (about 200 hertz) known as ripples, generated when multiple cells fire together within a narrow time



50 YEARS AGO

The Haunting of Borley Rectory —

This account of the evidence for abnormal happenings in what the late Harry Price described as “the most haunted house in England” well maintains the tradition of the Society for Psychical Research for honest and cautious study of alleged parapsychological phenomena. A heavy task was undertaken at the invitation of the Society by three trained investigators. Their story is at times as interesting as a detective novel; it reveals queer actions of some very curious people; it leaves very little to be explained of the actual haunt itself and a good deal to be puzzled over in the motives, actions and reactions of the people principally concerned. The general conclusion is that credulity, malobservation, trickery and fraud account for the great bulk of the recorded evidence.

From *Nature* 31 March 1956.

100 YEARS AGO

The System of the Stars. By Agnes M. Clerke —

There is much excellent sense in the French proverb, “Prends le premier conseil d’une femme, et non le second,” which expresses the view that the intuitive instinct of a woman is a safer guide to follow than her reasoning faculties; and although in these days it is considered ungracious to make this suggestion, evidence of its truth is not difficult to discover in most literary products of the feminine mind. It is no disparagement to Miss Clerke to say that even she shares this characteristic of her sex, so that sometimes she lets her sympathies limit her range of vision in the field of stellar research. No doubt this disposition is exercised unconsciously, but what is an attractive instinct when applied to ordinary affairs of life is derogatory when it influences the historiographic consideration of contributions to natural knowledge.

From *Nature* 29 March 1906.

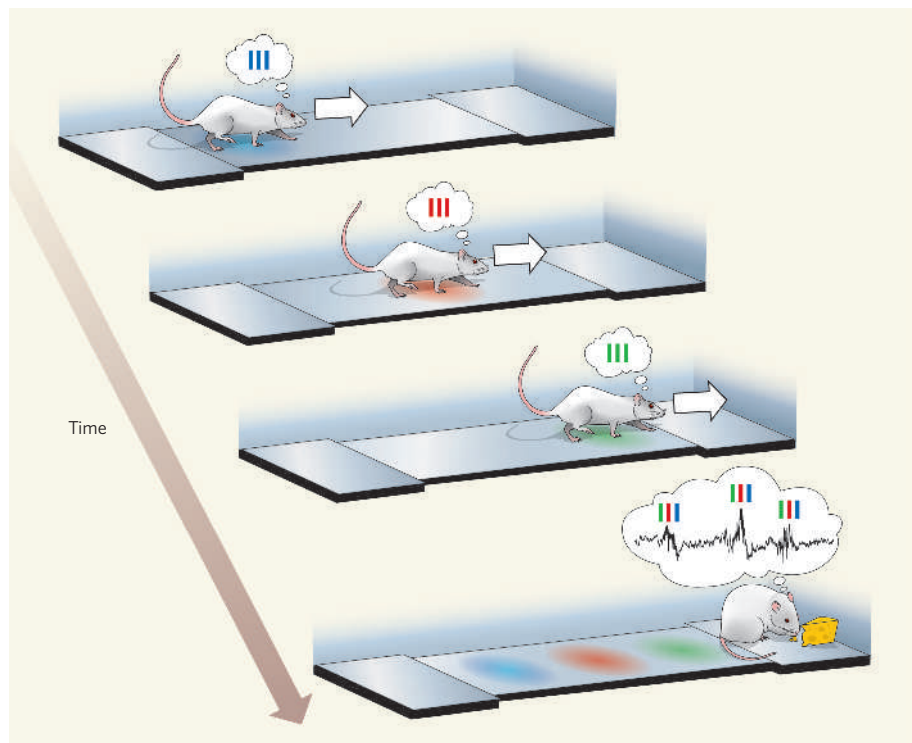


Figure 1 | Reverse replay. Three place cells (blue, red and green) in the hippocampus fire as a rat runs on a linear track. The coloured lines represent firing of place neurons. When the rat is rewarded with food at the end of the linear track, the hippocampus enters sharp-wave mode, and the firing sequences replay in reverse (that is, green, red, blue).

window⁷. Co-activation of interconnected neurons during ripples may result in long-lasting modifications of the synapses in the network (that is, the communication junctions between neurons)⁸.

Although reactivation during sleep may provide a mechanism for consolidation of recent memories, the mystery remains as to how memories can be maintained as distinct entities for hours or days in sleep-deprived subjects, considering that the participating neurons are probably involved in myriad events before the subject is finally allowed to take a nap. One clue comes from the observation that sharp waves occur also during waking states; for example, during resting, eating, drinking and brief breaks in exploration^{6,9}. Such ‘interleaved’ sharp waves may strengthen associations between recently activated cells only seconds after an event⁹.

Foster and Wilson² provide fascinating evidence for a mechanism that could generate such associations. They studied rats running back and forth on a narrow track, and they recorded neural activity from so-called place cells¹⁰. These hippocampal cells have spatial receptive fields, so each cell responds when the animal is in a particular location. Food was placed at the ends of the track, and the animals stopped after every lap to eat. When the rat paused, sharp waves emerged in its hippocampus. During these sharp waves, the place cells from the running period were reactivated, but their order of firing was reversed with respect to their earlier order of activation on the track (Fig. 1).

But how do neurons reverse firing sequences that were just stored in forward order? This might happen in at least two ways, one depending on the rat’s recent history and one reflecting its location in the environment. In the first possibility, the cells responding to place fields closest to the rest location are the first to reach the threshold for firing during the sharp wave because their synapses are still in a ‘facilitated’ state. Cells with fields that are farther away are less facilitated, so they take longer to reach the threshold. In the second option, cells fire in reverse order merely because firing probabilities of place cells increase with decreasing distance from the centre of their place fields, regardless of whether or not the rat has just passed through the fields. The latter possibility is partly ruled out because Foster and Wilson did not observe reverse replay in sharp waves recorded at the start of the session, before the rat began moving. This suggests that reverse reactivation is determined by the preceding sequence of events.

The million-dollar question, however, is what the brain gains by rewinding its neural record. At present, we do not know why sharp-wave-associated replay is forward in some circumstances (during sleep, say) and reversed in others. Foster and Wilson speculate that reverse replay has a role in reward-directed sequence learning during spatial navigation. Rewards (reinforcers) such as the food received at the ends of the track strengthen the preceding behavioural responses in a time-dependent manner such that the longer between the response and the reward, the less

the behaviour is strengthened^{11,12}. This mechanism is adaptive in evolutionary terms as it normally causes a fairly selective enhancement of those responses that generate the reward.

The authors hypothesize that the formation of associations between a reward and the representation of elements of a rat's trajectory in the immediate past is boosted during sharp-wave-associated replay by a neuromodulatory signal such as dopamine. Dopamine is a chemical released in the forebrain (in the striatum and cortex, and presumably the hippocampus) at the time of reward, especially when reward is not expected by the animal^{13–15}. Because ripple trains are variable in length, the effects of the boosting signal would be most reliable if it occurred at the beginning of the sharp wave; however, an early boost could be linked to the key later elements of the preceding firing sequence only if the sequence were reactivated in reverse order, as in Foster and Wilson's study. It remains to be seen

whether these speculations will stand up to experimental testing. At the moment, we do not know whether dopamine-releasing neurons fire in synchrony with hippocampal sharp waves.

If reverse replay is a mechanism for strengthening hippocampal sequence memories during goal-directed behaviour, several questions arise. For example, is the firing sequence stored as an ordered memory or as a unitary representation with a stronger representation of the later than of the earlier elements? Moreover, is reverse replay specific to sharp waves that coincide with reward? Sharp waves are observed during breaks without rewards. Do these sharp waves also exhibit reverse replay and, if so, are these associated with memory storage? Finally, can memories of events be stored without interleaved sharp waves? Whatever the answers may be, the discovery of reverse replay is bound to pave the way for more surprises. ■

Laura L. Colgin and Edvard I. Moser are at the Centre for the Biology of Memory, Norwegian University of Science and Technology, MTFs, 7489 Trondheim, Norway.

e-mail: edvard.moser@cbm.ntnu.no

1. Stickgold, R. *Nature* **437**, 1272–1278 (2005).
2. Foster, D. J. & Wilson, M. A. *Nature* **440**, 680–683 (2006).
3. Squire, L. R., Stark, C. E. & Clark, R. E. *Annu. Rev. Neurosci.* **27**, 279–306 (2004).
4. Wilson, M. A. & McNaughton, B. L. *Science* **265**, 676–679 (1994).
5. Lee, A. K. & Wilson, M. A. *Neuron* **36**, 1183–1194 (2002).
6. Buzsáki, G., Leung, L. W. & Vanderwolf, C. H. *Brain Res.* **287**, 139–171 (1983).
7. Klausberger, T. et al. *Nature* **421**, 844–848 (2003).
8. Buzsáki, G. *Neuroscience* **31**, 551–570 (1989).
9. O'Neill, J., Senior, T. & Csicsvari, J. *Neuron* **49**, 143–155 (2006).
10. O'Keefe, J. & Dostrovsky, J. *Brain Res.* **34**, 171–175 (1971).
11. Skinner, B. F. *The Behavior of Organisms* (Appleton-Century-Crofts, New York, 1938).
12. Hull, C. L. *A Behavior System* (Wiley, New York, 1952).
13. Berridge, K. C. & Robinson, T. E. *Brain Res. Rev.* **28**, 309–369 (1998).
14. Montague, P. R., Hyman, S. E. & Cohen, J. D. *Nature* **431**, 760–767 (2004).
15. Schultz, W. *Annu. Rev. Psychol.* **57**, 87–115 (2006).

QUANTUM METROLOGY

Size isn't everything

Samuel L. Braunstein

From probing living cells under a microscope to scanning the heavens for gravity waves, the limitations of precision measurements constrain our capacity to discover more about the world. But what exactly are those limits?

Just how accurate can measurements get? Whereas classical physics places no fundamental limits on how well we can do, in the quantum world it's a different story. Writing in *Physical Review Letters*, Giovannetti, Lloyd and Maccone¹ derive general limits for the precision with which a single variable can be measured quantum mechanically.

But is this new? After all, Heisenberg's uncertainty principle — one of the earliest results in quantum mechanics — already places a fundamental limitation on the precision with which we can make a measurement. In its simplest form, the uncertainty principle identifies so-called complementary observables, pairs of quantities for which knowing one quantity precisely means that the other can only be poorly known. This fundamental principle makes it impossible to learn everything about a quantum-mechanical system.

If we monitor only one quantity, however, there is no such in-principle limitation. In fact, this is exactly the strategy exploited in interferometric measurements, in which light travels down a pair of distinct paths and the difference between the two path lengths leads to an observable change in the output of the device. This path difference can be measured to an arbitrary accuracy. But what if we are given some constraint, such as a total energy budget or total light intensity? We all know

that it is easier to see in a well-lit room than in a dim one. Similarly, the higher the energy or light intensity in an interferometer, the higher its resolution. One may therefore ask, for a fixed budget, how small a path difference can be discerned?

Our intuition from everyday experience tells us that the most promising strategy for measuring a distance is to choose a measuring stick with marked intervals of length comparable to the distance we wish to measure. We would not, for example, choose a metre stick to measure a molecule. Following similar logic, we might choose the wavelength of light for our interferometer to be comparable to the path difference we want to measure. Surprisingly, Giovannetti and colleagues' latest result¹ can be used to show that, for optimal quantum strategies, there is no such bias to the size of our measuring stick or the separation of its tick marks.

An optimal strategy refers to a measurement procedure that minimizes the effects of noise on a signal. Ultimately, any measurement is limited by the amount of noise in the system: to discern a signal, the signal-to-noise ratio should be around one or larger. This premise underpins all parameter-estimation theory, both classical and quantum. Classically, statistical averaging over N repeated but independent measurements will lead to a $\sqrt{N}$ reduction in the noise. This improvement is known to

be optimal because it achieves the bound, known as the Cramér–Rao lower bound², that expresses the best accuracy that can be accomplished in the statistical estimation of a parameter. When this classical bound is generalized to repeated quantum measurements, the analogous quantum bound provides a tighter form of the uncertainty principle recast in the language of parameter estimation³. However, quantum theory allows much more freedom in choosing measurement strategies than is possible in the classical world.

One of the most bizarre features of the quantum world is quantum 'entanglement', which allows systems to exhibit stronger correlations than are possible classically. Using entanglement and other tricks, quantum mechanics has led us to devise sophisticated information-processing algorithms that one day may lie at the heart of the enormous speed-ups promised by quantum computation. For example, searching for a needle in a haystack would be much faster — in principle — on a quantum computer than a classical one. The possibility of using entangled systems and/or entangled measurements, and sophisticated algorithms built into measurement devices, raises questions about the ultimate (most general) quantum bounds to measurement.

Giovannetti and colleagues' key insight¹ into this question is to recast the measurement process in terms of quantum circuits, analogous to electrical circuits, with various quantum gates, similar to logic gates, representing different quantum-mechanical 'operators'. They then introduce black-box operators that perturb the quantum state in a known fashion, but by an unknown amount. Such an operation might, for instance, be adding a phase delay along one arm of an interferometer: the unknown parameter associated with the black box thus corresponds to the parameter we

would like to estimate. Once such a black box is conceptualized, it may be reused in the circuit again and again (each black box having the same unknown parameter). The beauty of this language lies in its generality, which allows a rich class of measurement strategies involving N such identical black boxes in a circuit of arbitrary design.

Using this formalism, Giovannetti *et al.* show that the optimal accuracy achievable in estimating the value of the black-box parameter can be obtained in a simple circuit with N black boxes, running on an N -fold entangled state. Surprisingly, recourse to entangled measurements (joint measurements of multiple paths of the circuit), or rearrangements of the circuit to correspond to sophisticated quantum-search strategies, will not lead to any further improvement.

What is this optimal performance? In fact, it depends entirely on the range of observable values of the black-box operator. In any circuit with N black boxes, the noise associated with the estimation of the black boxes' parameter will be reduced at most N -fold compared with the noise in the best circuit with only a single black box. That represents a considerable advantage over the $\sqrt{N}$ improvement of the classical case. The good (and reassuring) news is that this limit is exactly what one would have expected from a naive application of the good old Heisenberg uncertainty principle: it is none other than the Heisenberg limit.

So what relevance does all this have to the choice of size in our metre sticks? Well, let's return to our interferometer. For a given energy budget (or light intensity), but freedom in our choice of wavelength, we would naively expect the shorter wavelength to yield higher sensitivity. However, the longer the wavelength, the more photons we can squeeze into our interferometer. In other words, with the same budget, we can sample the black box exactly that many more times. Indeed, the Heisenberg-limited measurement is equally good, independent of our choice of measuring stick.

Two limitations to the strategy of Giovannetti *et al.*¹ lie in the quantum version of the Cramér–Rao bound on which it is based³. First, this bound can be reached only for problems involving single-parameter estimation, so extensions to multiple parameters may lead to different results. For instance, the estimation of the orientation of quantum spins (involving two unknown angles in three-dimensional space) can be enhanced by entangled measurements⁴. Second, the Cramér–Rao bound can be achieved only for an infinite number of repeated measurements. Thus, a result that expresses the approach to this asymptote would fill a gap in our current understanding. Indeed, it may be just this discrepancy that underlies the enhanced precision in determining the orientation of quantum spins using entangled measurements — an enhancement that vanishes in the

limit of an infinite number of spins⁴.

Currently, we are far from putting the ultimate bounds described by Giovannetti *et al.*¹ into practice. One example would be the Laser Interferometer Gravitational-Wave Observatory (LIGO), an exciting experiment that aims to detect tiny ripples in the fabric of space-time. The LIGO interferometer currently implements only classical strategies scaling as $1/\sqrt{N}$ (where N is the number of photons in the interferometer). In its current set-up, LIGO requires a circulating power of 10–20 kilowatts to achieve minimal sensitivities for detecting gravity waves. In principle, if we could implement a quantum-limited scheme, a similar sensitivity could be achieved with

only nanowatts. Such prospects promise an even brighter future for gravity-wave astronomy in the long term — and for precision measurement in general. ■

Samuel L. Braunstein is in the Department of Computer Science, University of York, York YO10 5DD, UK.

e-mail: schmuel@cs.york.ac.uk

1. Giovannetti, V., Lloyd, S. & Maccone, L. *Phys. Rev. Lett.* **96**, 010401 (2006).
2. Cramér, H. *Mathematical Methods of Statistics* 500–504 (Princeton Univ. Press, 1946).
3. Braunstein, S. L. & Caves, C. M. *Phys. Rev. Lett.* **72**, 3439–3442 (1994).
4. Gill, R. & Massar, S. *Phys. Rev. A* **61**, 042312 (2000).

MATERIALS SCIENCE

Nanostructures in a new league

John J. Rehr

Aperiodic materials do not surrender details of their structure as readily as do their crystalline counterparts. The latest computational solution to this problem brings aspects of 'the beautiful game' into play.

Investigations of crystalline materials through X-ray and neutron diffraction have been a triumph of experimental science, allowing structures ranging from complex minerals to proteins and DNA to be unravelled¹. But how can the structure of a material that is aperiodic — one that is non-crystalline, or cannot be crystallized — be determined? On page 655 of this issue², Juhás *et al.* present an intriguing solution to this question with a novel algorithm for reconstructing three-dimensional structures from 'pair distribution function' (PDF) data. Aperiodic materials are among the technologically most interesting nanoscale materials currently under study, and the approach could be widely applicable.

Several techniques exist for determining the local, atomic-scale structure of materials. These range from scanning tunnelling microscopy (STM) to spectroscopic methods that use X-rays, such as extended X-ray absorption fine structure (EXAFS) analysis. Each has its advantages and drawbacks. STM can give beautiful images, although not in three dimensions. For structural information to be inferred from spectroscopic techniques such as EXAFS, an accurate theoretical model relating spectra to structure is required³.

PDF analysis avoids some of these problems because it solely involves data on the distribution of distances between atoms in a structure — information that is readily obtained from X-ray or neutron-scattering experiments³. Why, then, is PDF not the method of choice for structure determination? The first factor is data quality: although the PDF technique has been known for decades, the lack of

high-resolution data has limited its applicability, as well as that of many other techniques. That situation is now changing with the latest generation of experiments using modern neutron and synchrotron X-ray sources.

The second crucial factor is that an algorithm must be found that solves the 'inverse problem'; that is, given a set of experimental data, how to extract the three-dimensional structure that must have created it. Determining the structure corresponding to PDF data, the question tackled by Juhás and colleagues², is just such an inverse problem. The inverse problem is usually not trivial, as it involves various assumptions about a material and, potentially, many material-dependent parameters. Solutions typically involve minimizing the mean squared deviation between the experimental data and the data predicted from a theoretical model of the structure. This process often needs significant computational resources, as it requires the 'direct problem' — that is, a theoretical model for the experimental signal resulting from a trial structure — to be solved many times in the process of finding the minimum.

Obtaining a solution to the inverse problem is equivalent to an optimization strategy for finding the global minimum of a quantity involving many variables among a forest of possible minima. Numerous advances have been made in such strategies, which are crucial in fields from economics to protein folding⁴. These include the development of 'genetic' algorithms inspired by the rules of evolutionary biology, and 'simulated annealing' techniques that mimic the way metals freeze into a

state of minimum energy. In the case of X-ray and neutron crystallography, the iterative 'shake-and-bake' algorithm¹ has been revolutionary. This method involves the random perturbation of the positions of atoms in a crystal until the lowest-energy state is found, and it has reduced the time required for determining crystal structures from months to just hours.

Juhás and colleagues call their approach² for inverting PDF data the 'Liga algorithm', because the method is modelled on the rules of promotion and relegation that determine the position of participating teams in most of the world's soccer leagues. Teams correspond to trial clusters of atoms; 'winning' clusters (those with the smallest errors between the model and the experiment) are iteratively promoted, whereas losing ones (those with the largest errors) are relegated, so that an optimal global structure is more quickly found. The authors show that their algorithm can determine a number of nanoscale structures, such as that of the 'buckyball' C₆₀ molecule, with a perfect success rate. Genetic algorithms, in contrast, take considerably longer and have far lower rates of success.

So what are the limits of this approach, and can it be extended to other global-optimization problems? The limits are typically reached when there are more parameters in the theoretical model than can be represented by data, so that the inverse problem becomes 'ill-conditioned' — that is, it has unstable solutions. Optimization strategies must therefore include some way of stabilizing the solutions. Some of these approaches, such as choosing model parameters by guesswork, can involve more than a whiff of the black art, and potentially produce results that vary widely from one investigator to the next.

Alternative methods using powerful statistical methods such as bayesian analysis have been developed, which can avoid the arbitrariness of choosing model parameters⁵. They achieve stability by taking into account a priori information in order to constrain the overall probability distribution for a particular structure. Strategies such as the Liga algorithm could be extended significantly by including known structural information based on a system's physical and chemical properties or knowledge derived from theory and computational materials science. It may well then be possible to resolve heterogeneous nanostructures containing many hundreds of atoms. ■

John J. Rehr is in the Department of Physics, University of Washington, Seattle, Washington 98195-1560, USA.

e-mail: jjr@phys.washington.edu

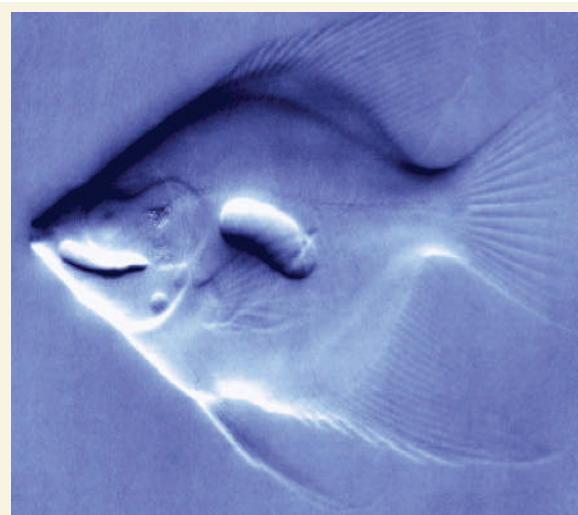
1. Miller, R. *et al.* *Science* **259**, 1430–1433 (1993).
2. Juhás, P., Cherba, D. M., Duxbury, P. M., Punch, W. F. & Billinge, S. J. L. *Nature* **440**, 655–658 (2006).
3. Rehr, J. J. & Albers, R. *Rev. Mod. Phys.* **72**, 621–654 (2000).
4. Wales, D. J. & Scheraga, H. A. *Science* **285**, 1368–1372 (1999).
5. Krappe, H. J. & Rossner, H. H. *Phys. Rev. B* **66**, 184303–184323 (2002).

X-RAY IMAGING

Soft focus

From Wilhelm Conrad Röntgen's first snapshot of his wife's hand in 1895, to the security scanner that blows Arnold Schwarzenegger's cover in *Total Recall*, the use of X-rays to image dense objects has been part of common lore. Franz Pfeiffer and colleagues (*Nature Phys.* doi:10.1038/nphys265; 2006) now realign the popular view. They use X-rays to generate high-contrast images not only of bone, but also of the soft tissues that surround them. The approach could readily be used to improve the diagnostic power of existing medical-imaging equipment.

Conventional medical X-ray imaging uses the fact that the harder and denser the body tissues are, the more radiation they absorb, and the more contrast they produce on X-ray films. This makes it easy to distinguish bones and other dense bodies, such as tumours, from surrounding tissues. But discerning details of softer tissues from only the contrast in absorption is difficult.



When an X-ray passes through tissue, however, it is not just absorbed: its phase is changed too. And this phase shift is more sensitive to variations in the composition of soft-tissue structures than is absorption. But until now, extracting information about phase has required interferometric reconstruction techniques and bathing the target object in the ultra-high-intensity radiation of a synchrotron particle accelerator.

Pfeiffer *et al.* use a sequence of phase-contrast gratings to

manipulate the relative phases of the X-rays that illuminate and subsequently emerge from an object. They can thus generate phase-contrast images — for example this 50-mm × 50-mm picture of an angelfish — using commercial X-ray sources at much lower intensity, and cost, than has previously been possible. The authors note that, as well as improving the detail in X-ray images, their approach could be adapted for use with other low-intensity radiation sources, such as neutrons and ions.

Ed Gerstner

COGNITIVE SCIENCE

Brain development and IQ

Richard Passingham

If intelligence is partly determined by our genes, how does brain development relate to IQ? An attempt to answer this question measures the size of the outer layer of the brain, the cortex, with surprising results.

Shaw and colleagues (page 676 of this issue)¹ have investigated whether there is a relationship between intelligence and physical dimensions of the brain. Specifically, they measure the thickness of the cortex; the complex computations carried out by the brain depend on the firing of the cortical cells. The authors' results indicate that intelligence can be related to how the cortex changes during development.

Rather than making structural measurements in post-mortem brains, Shaw and colleagues used magnetic resonance imaging (MRI) in living subjects. This allowed the authors to obtain images from people whose IQ could also be tested so as to look for correlations between the two measures. Moreover, detecting anatomical features associated with an

individual's intelligence requires a large pool of subjects, because any effects may be small and could be missed if the sample size is inadequate. The use of imaging, rather than post-mortem measurements, allows data to be gathered from a sufficient number of individuals.

The authors scanned 307 children from the age of six years and followed them through adolescence with further scans. For each child, the authors estimated intelligence using subtests of the Wechsler Intelligence Scales — the most commonly used IQ tests. An alternative approach would have been to look at a cross-sectional sample of children and adolescents of different ages, scanned only once each. But, as the authors note, such methods are open to many objections: for example, teaching

practices may change over time, which would affect the IQ scores.

Shaw and colleagues find no significant correlation between cortical thickness and intelligence in their data from young children. Yet they cite a study of adults by McDaniel² that reports a modest correlation of 0.3 between intelligence and the total volume of the brain. The reason for the different results could be that the relevant factor is the total area of the cortex rather than its thickness, but it turns out that this is probably not the case. As the children were followed up, the nature of the relationship changed. In young children, the correlation tended to be negative, but in late childhood, around the age of ten, it was positive.

The authors illustrate this point by plotting continuous curves of cortical thickness for subjects from the ages of seven to nineteen, dividing the sample into three groups on the basis of their scores in the IQ tests: those of 'superior', 'high' and 'average' intelligence. IQ measures are normalized to the age group, and should in theory remain the same as the children age. Figure 2 on page 677 shows the curves for cortical thickness in brain areas that show different developmental patterns according to intelligence. Children in the group with superior intelligence have a thinner cortex in these areas in early childhood, but cortical thickness increases sharply until age eleven compared with the other groups, before decreasing through adolescence. The authors note that those of superior intelligence show a prolonged period of prefrontal cortical gain and the most rapid rate of change.

These differential changes do not occur in all cortical areas. The most notable positive correlations with IQ in late childhood occur in the prefrontal cortex. This region lies at the top of the information-processing hierarchy,

receiving highly processed information from all five senses³. The brain areas showing the biggest difference in the shape of the growth curve between those with superior intelligence and the other groups lie in the lateral and medial frontal gyri. But are these the areas that are most active when subjects perform IQ tests? This aspect can be assessed by functional MRI, which provides an indirect measure of the increase in arterial blood flow to areas in which cellular activity is increased. Previously, subjects have been scanned while taking non-verbal tests that measure IQ, and increased activity has been found in the lateral and medial prefrontal cortex — regions that are among those highlighted by Shaw and colleagues' developmental measures^{4,5}. Furthermore, individual differences in IQ are correlated with the amplitude of the functional MRI signal in the lateral prefrontal cortex⁶.

We know that variations in general intelligence, or *g*, among people depend to a great extent on genetic differences⁷. So, if *g* is highly heritable and the increase in the thickness of the prefrontal cortex is related to *g*, it is tempting to assume that this developmental change in brain structure is determined by a person's genes. But one should be very wary of such a conclusion. The body's development is intimately linked to interactions with its environment. For example, in a classic experiment, Rosenzweig and Bennett⁸ showed that the thickness of the cortex in adult rats is affected by the degree to which the animals' early environment is enriched in terms of activities. Even in human adults, structural changes can be seen in the cortical grey matter as a result of practice⁹. Thus, it could be that people with superior intelligence also live in a richer social and linguistic environment, and that it is this that accounts for the sharp increase

in the thickness of their prefrontal cortex in late childhood. However, Thompson and colleagues¹⁰ previously looked for genetic influences on brain structure by comparing the cortical thickness of pairs of identical and non-identical twins. They found that some regions, including the frontal cortex, are, to use their words, under "tight genetic control".

Shaw and colleagues speculate that differences in the shape of the growth curves of cortical thickness could be influenced by various factors. These include the number of neurons that collect in the subplate under the cortex during late fetal development, the development of the myelin sheath that insulates the fibres of the neurons, and the selective elimination at puberty of neuronal connections that are not useful. Testing these hypotheses will require animal experiments that measure cellular development. Studies in animals have the advantage that the relative influence of genetics and experience can be disentangled, and so should provide a clearer picture of how intellectual ability is affected by the factors that underpin cortical development. ■

Richard Passingham is in the Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford OX1 3UD, UK.
e-mail: dick.passingham@psy.ox.ac.uk

1. Shaw, P. *et al.* *Nature* **440**, 676–679 (2006).
2. McDaniel, M. *Intelligence* **33**, 337–346 (2005).
3. Passingham, R. E., Rowe, J. B. & Sakai, K. in *Attention in Action* (eds Humphreys, G. W. & Riddoch, M. J.) 263–286 (Psychology Press, Hove, 2005).
4. Duncan, J. *et al.* *Science* **289**, 457–460 (2000).
5. Christoff, K. *et al.* *NeuroImage* **14**, 1136–1149 (2001).
6. Gray, J. R., Chabris, C. F. & Braver, T. S. *Nature Neurosci.* **6**, 316–322 (2003).
7. Plomin, R. & Spinath, F. M. *Trends Cogn. Sci.* **6**, 169–176 (2002).
8. Rosenzweig, M. R. & Bennett, E. L. *Behav. Brain Res.* **78**, 57–65 (1996).
9. Draganski, B. *et al.* *Nature* **427**, 311–312 (2004).
10. Thompson, P. M. *et al.* *Nature Neurosci.* **4**, 1253–1258 (2001).

ENVIRONMENTAL CHEMISTRY

Boiling up an acid plume

There is more than just a sizzle when red-hot lava meets the sea. The plumes seen in this picture consist not only of steam produced by the evaporation of water, but also of aerosols and gases that stem from the reaction between the lava and salt water.

M. Edmonds and T. M. Gerlach have investigated the composition of such plumes produced by lava from Kilauea Volcano, Hawaii (*Earth Planet. Sci. Lett.* doi: 10.1016/j.epsl.2006.02.005). Their main tool was open-path Fourier transform

infrared spectroscopy, which allowed remote sensing of the plumes and estimation of the amounts of various components — water, carbon dioxide, nitrogen dioxide, sulphur dioxide and hydrogen chloride. The most notable of Edmonds and Gerlach's conclusions stem from their analyses of this last species, HCl.

First, from thermodynamic considerations they calculate that the HCl gas is created following the hydrolysis of magnesium chloride salts (and not of sodium chloride, as an

alternative explanation has it). Second, given that conclusion, they estimate how much HCl is produced by the lava-seawater interaction. The outcome depends on various assumptions and factors, including the type, extent and duration of the lava flow.

Edmonds and Gerlach estimate that a lava flow of $1 \text{ m}^3 \text{ s}^{-1}$ could in principle produce 3.7 kg s^{-1} of HCl, or 300 tonnes daily. For various reasons that they discuss, this number is likely to be much lower (3–30 tonnes). Figures



of this latter order of magnitude produce only localized high concentrations of HCl gas and acid rain. But the authors point out that in the past the story must have had a more serious edge. Eruptions of Hawaiian volcanoes in 1840, 1919

and 1950 produced massive lava flows, with sustained lava fluxes entering the sea. The result was an estimated HCl output of 200–2,200 tonnes per day over several weeks, a much more serious environmental hazard.

Tim Lincoln

BRIEF COMMUNICATIONS

Shape memory in spider draglines

The torsional properties of spider silk add to its list of remarkable physical credentials.

The ductility and strength of spider draglines means that they outperform the best synthetic fibres^{1–5}, but surprisingly little is known about the torsional properties of this remarkable filament. Unlike a mountain climber swinging from a rope, a spider suspended from its silk thread hardly ever twists. Here we show that a spider dragline has a torsional shape ‘memory’ in that it can reversibly and totally recover its initial form without any external stimulus; its observed relaxation dynamics indicate that these biological molecules have successively different torsional constants.

The torsion pendulum is the best tool for exploring the twist properties of threads because its period of oscillation does not depend on the amplitude of oscillation. In our experimental set-up, we used a small plastic or copper rod to represent the weight of the spider and suspended it from a 10-cm thread. The ‘spider’ rod was turned through about 90° and the dynamical responses of different suspending threads were registered by a camera linked to a computer.

When a thread of the synthetic organic polymer Kevlar is twisted from equilibrium, the rod oscillates gently around its original position. In this case, the response of the filament is elastic, with little energy dissipation (Fig. 1a): its oscillating behaviour is typical of a weakly damped, torsional pendulum.

The torsional stability of a copper thread is greater after being twisted from its original rest position — the rod barely oscillates around its new (deformed) equilibrium position (Fig. 1b). The soft metallic copper thread, unlike the tough Kevlar filament, displays the strong damping typical of high energy dissipation⁶. The mechanical properties of the copper thread confer good torsional stability — its rigidity increases because of work-hardening due to dislocations, stopping it from easily untwisting. In fact, copper becomes brittle after few twists⁷.

Unlike either of these filaments, dragline silk from the spider *Araneus diadematus* (Fig. 2), when turned from its initial position, scarcely oscillates around the new pseudo-equilibrium position because of its high damping coefficient (Fig. 1c, inset). This response is independent of friction arising from resistance to the air. Unlike copper thread, spider silk retains its torsional qualities through several cycles. Also, the spider silk completely recovers its original position in what resembles an exponential manner (Fig. 1c). This requires the filament to exhibit a series of different torsional constants. The response sets spider silk apart from either Kevlar or copper.

We compared spider silk’s response with that of a synthetic thread of ‘shape-memorizing’ nickel–titanium alloy (Nitinol), which is widely used in industry⁷. Like the spider dragline, the excited Nitinol thread weakly oscillates around an equilibrium position that differs from the original one (Fig. 1d). Its properties are due to a coherent diffusionless deformation of the shape-memory crystal without any dislocation⁸. However, Nitinol has to be



Figure 2 | A spider's silk. **a**, Scanning electron micrograph of a dragline. Scale bar, 10 μm . **b**, The thread's unusual torsional properties prevent an abseiling spider from swinging, a movement that might attract predators.

heated to 90 °C to recover its original shape. By contrast, the spider has evolved a shape-memory material that needs no external stimulus for total recovery, even though it takes time.

Our results show that spider-dragline silk, in addition to its outstanding tensile strength and toughness, has unrivalled torsional qualities that stop the spider twisting and swinging. In view of the different torsional relaxations underlying the behaviour shown in Fig. 1c, testing the relaxation of spider-silk analogue proteins such as poly-L-alanine and poly-L-glycine could be warranted — to see whether reconstruction of sacrificial bonds forms the basis for the observed behaviour of silk. Tiny torque effects in other biological molecules such as DNA^{9,10} also wait to be unravelled.

Olivier Emile*, Albert Le Floch*, Fritz Vollrath†

*Laboratoire de Physique des Lasers, UMR CNRS 6627, Université de Rennes 1, 35042 Rennes, France
e-mail: olivier.emile@univ-rennes1.fr

†Department of Zoology, South Parks Road, Oxford OX1 3PS, UK

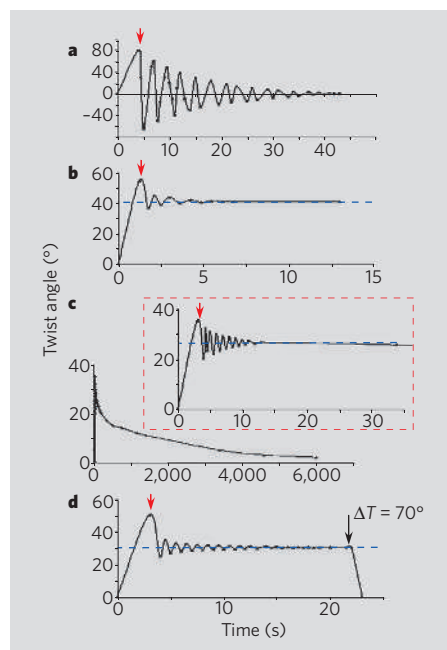


Figure 1 | Relaxation dynamics of a torsion pendulum for four different threads. Red arrows indicate the end of the excitation and the start of the free relaxation period. Blue dotted line, new equilibrium position. **a**, Kevlar thread (poly-*p*-phenylene terephthalamide) of diameter 10 μm ; the suspended mass is 0.5 g and the torsional constant C is $10^{-1} \text{ N m rad}^{-1}$. **b**, Copper thread of diameter 50 μm ; the suspended mass is 2 g. **c**, Thread of the spider *Araneus diadematus*, drawn from a falling spider and handled carefully to avoid stretching; the suspended mass is 0.1 g. Inset: zoom of the start of the self-relaxation dynamics. **d**, Shape-memory Nitinol alloy thread of diameter 100 μm ; the suspended mass is 3 g. The response of this thread is similar to that of spider dragline, but the alloy needs to be heated from 20 °C to 90 °C in order to relax to its original shape.

- Gosline, J. M., Denny, M. W. & DeMont, M. E. *Nature* **309**, 551–552 (1984).
- Lazaris, A. et al. *Science* **295**, 472–476 (2002).
- Jin, H.-J. & Kaplan, D. L. *Nature* **424**, 1057–1061 (2003).
- Craig, C. L. *Spiderwebs and Silk* (Oxford Univ. Press, New York, 2003).
- Becker, N. et al. *Nature Mater.* **2**, 278–283 (2003).
- Feynman, R. P., Leighton, R. B. & Sands, M. *The Feynman Lectures on Physics Vol. 2* (Addison-Wiley, London, 1964).
- Ashley, S. *Sci. Am.* **289**, 84–91 (2003).
- Bhattacharya, K., Conti, S., Zanzotto, G. & Zimmer, J. *Nature* **428**, 55–59 (2004).
- Bryant, Z. et al. *Nature* **424**, 338–341 (2003).
- Koster, D. A., Croquette, V., Dekker, C., Shuman, S. & Dekker, N. H. *Nature* **434**, 671–674 (2005).

Received 17 October 2005; accepted 20 February 2006.
Competing financial interests: declared none.
doi:10.1038/440621a

REVIEWS

Eukaryotic evolution, changes and challenges

T. Martin Embley¹ & William Martin²

The idea that some eukaryotes primitively lacked mitochondria and were true intermediates in the prokaryote-to-eukaryote transition was an exciting prospect. It spawned major advances in understanding anaerobic and parasitic eukaryotes and those with previously overlooked mitochondria. But the evolutionary gap between prokaryotes and eukaryotes is now deeper, and the nature of the host that acquired the mitochondrion more obscure, than ever before.

New findings have profoundly changed the ways in which we view early eukaryotic evolution, the composition of major groups, and the relationships among them. The changes have been driven by a flood of sequence data combined with improved—but by no means consummate—computational methods of phylogenetic inference. Various lineages of oxygen-shunning or parasitic eukaryotes were once thought to lack mitochondria and to have diverged before the mitochondrial endosymbiotic event. Such key lineages, which are salient to traditional concepts about eukaryote evolution, include the diplomonads (for example, *Giardia*), trichomonads (for example, *Trichomonas*) and microsporidia (for example, *Vairimorpha*). From today's perspective, many key groups have been regrouped in unexpected ways, and aerobic and anaerobic eukaryotes intermingle throughout the unfolding tree. Mitochondria in previously unknown biochemical manifestations seem to be universal among eukaryotes, modifying our views about the nature of the earliest eukaryotic cells and testifying to the importance of endosymbiosis in eukaryotic evolution. These advances have freed the field to consider new hypotheses for eukaryogenesis and to weigh these, and earlier theories, against the molecular record preserved in genomes. Newer findings even call into question the very notion of a 'tree' as an adequate metaphor to describe the relationships among genomes. Placing eukaryotic evolution within a time frame and ancient ecological context is still problematic owing to the vagaries of the molecular clock and the paucity of Proterozoic fossil eukaryotes that can be clearly assigned to contemporary groups. Although the broader contours of the eukaryote phylogenetic tree are emerging from genomic studies, the details of its deepest branches, and its root, remain uncertain.

The universal tree and early-branching eukaryotic lineages

The universal tree based on small-subunit (SSU) ribosomal RNA¹ provided a first overarching view of the relationships between the different types of cellular life. The relationships among eukaryotes recovered from rRNA², backed up by trees of translation elongation factor (EF) proteins³, provided what seemed to be a consistent, and hence compelling, picture (Fig. 1). The three protozoa at the base of these trees (*Giardia*, *Trichomonas* and *Vairimorpha*), along with *Entamoeba* and its relatives, were seen as members of an ultrastructurally simple, paraphyletic group of eukaryotes called the Archezoa⁴. Archezoa were thought to primitively lack mitochondria, having split from the main trunk of the eukaryotic tree before the mitochondrial endosymbiosis: all other eukaryotes contain mitochondria because

they diverged after this singular symbiotic event⁵. Therefore, Archezoa were interpreted as contemporary descendants of a phagotrophic, nucleated, amitochondriate cell lineage that included the host for the mitochondrial endosymbiont⁶. The apparent agreement between molecules and morphology depicted the relative timing of the mitochondrial endosymbiosis (Fig. 1) as a crucial, but not ancestral, event in eukaryote phylogeny.

Chinks in the consensus

Mitochondrial genomes studied so far encode less than 70 of the proteins that mitochondria need to function⁵; most mitochondrial proteins are encoded by the nuclear genome and are targeted to

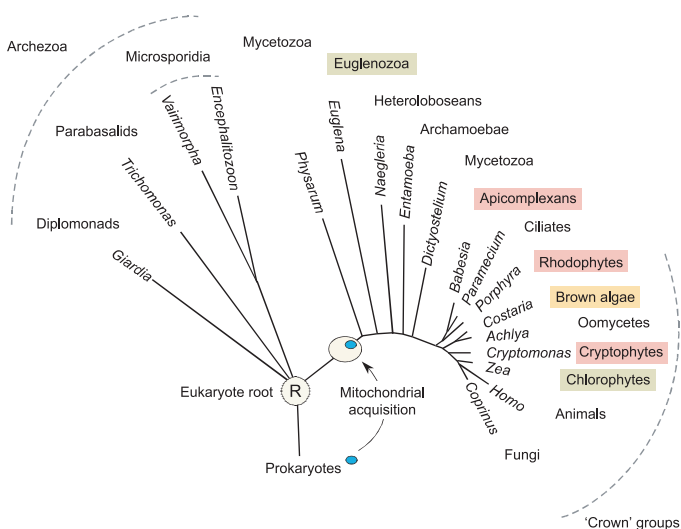


Figure 1 | The general outline of eukaryote evolution provided by rooted rRNA trees. The tree has been redrawn and modified from ref. 92. Until recently, lineages branching near the root were thought to primitively lack mitochondria and were termed Archezoa⁴. Exactly which archezoans branched first is not clearly resolved by rRNA data², hence the polytomy (more than two branches from the same node) involving diplomonads, parabasalids and microsporidia at the root. Plastid-bearing lineages are indicated in colours approximating their respective pigmentation. Lineages furthest away from the root, including those with multicellularity, were thought to be the latest-branching forms and were sometimes misleadingly (see ref. 60) called the 'crown' groups.

¹School of Biology, The Devonshire Building, University of Newcastle upon Tyne, Newcastle NE1 7RU, UK. ²Institute of Botany III, University of Düsseldorf, D-40225 Düsseldorf, Germany.

mitochondria using a protein import machinery that is specific to this organelle⁷. The mitochondrial endosymbiont is thought to have belonged to the α -proteobacteria, because some genes and proteins still encoded by the mitochondrial genome branch in molecular trees among homologues from this group^{5,8}. Some mitochondrial proteins, such as the 60- and 70-kDa heat shock proteins (Hsp60, Hsp70), also branch among α -proteobacterial homologues, but the genes are encoded by the host nuclear genome. This is readily explained by a corollary to endosymbiotic theory called endosymbiotic gene transfer⁹: during the course of mitochondrial genome reduction, genes were transferred from the endosymbiont's genome to the host's chromosomes, but the encoded proteins were re-imported into the organelle where they originally functioned. With the caveat that gene origin and protein localization do not always correspond⁹, any nuclear-encoded protein that functions in mitochondria and clusters with α -proteobacterial homologues is most simply explained as originating from the mitochondrion in this manner.

By that reasoning¹⁰, the discovery of mitochondrial Hsp60 in *E. histolytica* was taken as evidence that its ancestors harboured mitochondria. A flood of similar reports on mitochondrial Hsp60 and Hsp70 from all key groups of Archezoa ensued¹¹, suggesting that

their common ancestor also contained mitochondria. At face value, those findings falsified the central prediction of the archezoan concept. However, suggestions were offered that lateral gene transfer (LGT) in a context not involving mitochondria could also account for the data. But that explanation, apart from being convoluted, now seems unnecessary: the organisms once named Archezoa for lack of mitochondria not only have mitochondrial-derived proteins, they have the corresponding double-membrane-bounded organelles as well.

Mitochondria in multiple guises

The former archezoans are mostly anaerobes, avoiding all but a trace of oxygen, and like many anaerobes, including various ciliates and fungi that were never grouped within the Archezoa, they are now known to harbour derived mitochondrial organelles—hydrogenosomes and mitosomes. These organelles all share one or more traits in common with mitochondria (Fig. 2), but no traits common to them all, apart from the double membrane and conserved mechanisms of protein import, have been identified so far. Mitochondria typically—but not always (the *Cryptosporidium* mitochondrion lacks DNA¹²)—possess a genome that encodes components involved in oxidative phosphorylation⁵. With one notable exception¹³, all hydrogenosomes

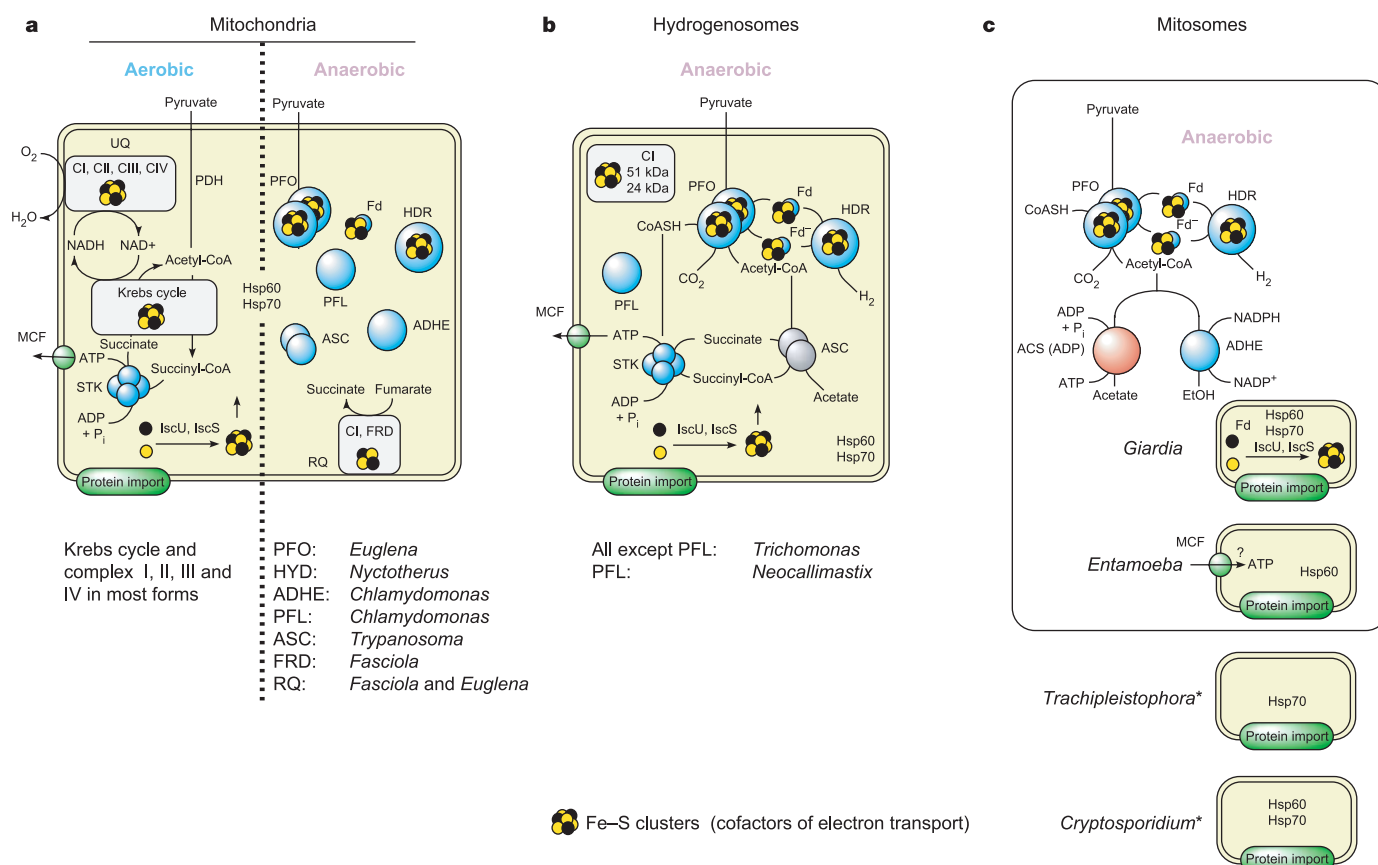


Figure 2 | Enzymes and pathways found in various manifestations of mitochondria. Proteins sharing more sequence similarity to eubacterial than to archaeobacterial homologues are shaded blue; those with converse similarity pattern are shaded red; those whose presence is based only on biochemical evidence are shaded grey; those lacking clearly homologous counterparts in prokaryotes are shaded green. **a**, Schematic summary of salient biochemical functions in mitochondria^{5,88}, including some anaerobic forms^{16,17}. **b**, Schematic summary of salient biochemical functions in hydrogenosomes^{14,19}. **c**, Schematic summary of available findings for mitosomes and 'remnant' mitochondria^{32–34,93}. The asterisk next to the *Trachipleistophora* and *Cryptosporidium* mitosomes denotes that these organisms are not anaerobes in the sense that they do not inhabit O₂-poor

niches, but that their ATP supply is apparently O₂-independent. UQ, ubiquinone; CI, mitochondrial complex I (and II, III and IV, respectively); NAD, nicotinamide adenine dinucleotide; MCF, mitochondrial carrier family protein transporting ADP and ATP; STK, succinate thiokinase; PFO, pyruvate:ferredoxin oxidoreductase; PDH, pyruvate dehydrogenase; CoA, coenzyme A; Fd, ferredoxin; HDR, iron-only hydrogenase; PFL, pyruvate:formate lyase; ASC, acetate-succinate CoA transferase; ADHE, bi-functional alcohol acetaldehyde dehydrogenase; FRD, fumarate reductase; RQ, rhodoquinone; Hsp, heat shock protein; IscU, iron-sulphur cluster assembly scaffold protein; IscS, cysteine desulphurase; ACS (ADP), acetyl-CoA synthase (ADP-forming).

and mitochondria studied so far lack a genome. The organisms in which they have been studied generate ATP by fermentations involving substrate-level phosphorylations, rather than through chemiosmosis involving an F_1/F_0 -type ATPase^{12,14,15}. *Entamoeba*, *Giardia* and *Trichomonas* live in habitats too oxygen-poor to support aerobic respiration¹⁴, while others, like *Cryptosporidium* and microsporidia have drastically reduced their metabolic capacities during adaptation to their lifestyles as intracellular parasites^{12,15}.

Between aerobic mitochondria, which use oxygen as the terminal electron acceptor of ATP-producing oxidations, and *Nyctotherus* hydrogenosomes, which (while retaining a mitochondrial genome) use protons instead of oxygen¹³, there are a variety of other anaerobically functioning mitochondria. They occur in protists such as *Euglena*, but also in multicellular animals such as *Fasciola* and *Ascaris*, which typically excrete acetate, propionate or succinate, instead of H_2O or H_2 , as their major metabolic end-products^{16,17}. Hence, mitochondria, hydrogenosomes and mitosomes are viewed most simply as variations on a single theme, one that fits neatly within the framework provided by classical evolutionary theory¹⁸. They are evolutionary homologues that share similarities because of common ancestry, but—like forelimbs in vertebrates—differ substantially in form and function across lineages owing to descent with modification.

Hydrogen-producing mitochondria

Hydrogenosomes oxidize pyruvate to H_2 , CO_2 and acetate, making ATP by substrate-level phosphorylation¹⁹ that they export to the cytosol using a mitochondrial-type ADP/ATP carrier^{20,21}. They have been identified in trichomonads, chytridiomycetes and ciliates^{13,22}; their hydrogen excretion helps to maintain redox balance¹⁴ in these organisms. Important similarities between *Trichomonas* hydrogenosomes and mitochondria include the use of common protein import pathways²³, conserved mechanisms of iron–sulphur-cluster assembly²⁴, conserved mechanisms of NAD^+ regeneration²⁵, and conservation of a canonical ATP-producing enzyme of the mitochondrial Krebs cycle—succinate thiokinase²⁶. On the basis of electron microscopy and ecology, additional, and diverse, eukaryotic lineages are currently suspected to contain hydrogenosomes^{27,28}, but hydrogen production—the defining characteristic of hydrogenosomes¹⁹—by those organelles has not yet been shown.

In contrast to most mitochondria, hydrogenosomes typically contain pyruvate:ferredoxin oxidoreductase (PFO) and iron [Fe] hydrogenase. Common among anaerobic bacteria, these enzymes prompted the early suggestion that trichomonad hydrogenosomes arose from a *Clostridium*-like endosymbiont²⁹. In a recent rekindling of that idea^{30,31}, trichomonad hydrogenosomes were suggested to be hybrid organelles, derived from an endosymbiotic anaerobic bacterium (the source of PFO and hydrogenase genes), a failed mitochondrial endosymbiosis (the source of nuclear genes for mitochondrial Hsp60 and Hsp70), plus LGT from a mitochondrially related (but non-mitochondrial) donor (the source of NADH dehydrogenase). However, independent work suggested a mitochondrial, rather than hybrid, origin of the *Trichomonas* NADH dehydrogenase²⁵. Furthermore, the hybrid hypothesis fails to account for the presence of [Fe]-hydrogenase homologues in algal chloroplasts, PFO homologues in *Euglena* mitochondria, or the presence of either enzyme and hydrogenosomes in other eukaryotic lineages²⁵; hence, a single common ancestry of mitochondria and hydrogenosomes sufficiently accounts for current observations.

Mitochondria reduced to bare bones

Mitosomes were discovered in *Entamoeba*³² as mitochondrion-derived organelles that have undergone more evolutionary reduction than hydrogenosomes. They are also found in *Giardia*³³ and microsporidia³⁴. Mitosomes seem to have no direct role in ATP synthesis because, so far, they have been found only among eukaryotes whose core ATP synthesis occurs in the cytosol¹⁴ or among energy

parasites¹⁵. Mitosomes import proteins in a mitochondrial-like manner^{35–37}, and *Giardia* mitosomes contain two mitochondrial proteins of Fe–S cluster assembly—cysteine desulphurase (IscS) and iron-binding protein (IscU)³³. Fe–S clusters are essential for life: they are cofactors of electron transfer, catalysis, redox sensing and ribosome biogenesis in eukaryotes³⁸. Fe–S cluster assembly is an essential function of yeast mitochondria³⁸ and it has been widely touted as a potential common function for mitochondrial homologues^{15,22}. It is the only known function of *Giardia* mitosomes, which, like *Trichomonas* hydrogenosomes^{24,37}, promote assembly of [2Fe–2S] clusters into apoferredoxin *in vitro*³³. By contrast, and (so far) uniquely among eukaryotes, *Entamoeba* uses two proteins of non-mitochondrial ancestry for Fe–S cluster assembly³⁹; the location of this pathway in *Entamoeba* is currently unknown.

Branch migrations and evolutionary models

The discovery of mitochondrial homologues in *Giardia*, *Trichomonas* and microsporidians, which had been the best candidates for eukaryotes that primitively lacked mitochondria, has pinned the timing of the mitochondrial origin to the ancestor of all eukaryotes studied so far. But that does not mean that the basal position of these groups in the SSU rRNA tree (Fig. 1) and EF trees³ is necessarily incorrect. That issue hinges on efforts to construct reliable rooted phylogenetic trees depicting ancient eukaryotic relationships: a developing area of research that is fraught with difficulties. The tempo and mode of sequence evolution is far more complicated than is assumed by current mathematical models that are used to make phylogenetic trees⁴⁰. In computer simulations, where the true tree is known, model mis-specification can produce the wrong tree with strong support⁴¹.

Different sites in molecular sequences evolve at different rates, and failure to accommodate this rate variation, something early methods failed to do, can lead to strongly supported but incorrect trees owing to a common problem called ‘long-branch-attraction’⁴². This occurs when branches that are long or ‘fast evolving’, relative to others in the tree, cluster together irrespective of evolutionary relationships. The molecular sequences of *Giardia*, *Trichomonas* and microsporidia often form long branches in trees and thus are particularly prone to this problem^{25,43,44}. The traditional models that placed microsporidia deep within trees^{2,3} assumed that all sequence sites evolved at the same rate, even though they clearly do not. In these trees, the long-branch microsporidia are next to the long branches of the prokaryotic outgroups. More data and better models have produced trees that agree in placing microsporidia with fungi^{45,46}, suggesting that the deep position of microsporidia in early trees was indeed an artefact.

The position of *Giardia* and *Trichomonas* sequences at the base of eukaryotic molecular trees is also suspect, given that they also form long branches in the trees that place them in this way, and because other trees and models place them together as an internal branch of a rooted eukaryotic tree⁴⁷. Resolving which position is correct is particularly important, because *Giardia* and *Trichomonas* are still commonly referred to as ‘early-branching’ eukaryotes. Given the evident uncertainties of such phylogenies, and the importance of the problem, the onus is on those who would persist in calling these species ‘early branching’ to show that trees placing them deep explain the data significantly better than trees that do not.

The root of the eukaryotic tree

The usual way to root a phylogenetic tree is by reference to an outgroup; the rRNA and EF trees used prokaryotic sequences to root eukaryotes on either the *Giardia*, *Trichomonas* or microsporidia branch (Fig. 1), but these rootings have not proved robust^{43–45}. The sequences of outgroups are often highly divergent compared to those of the ingroup, making it difficult to avoid model mis-specification and long-branch-attraction^{44,48}.

An alternative method of rooting an existing tree is to look for rare

changes in a complex molecular character where the ancestral state can be inferred. This method was used⁴⁹ to infer that the root of the eukaryotic tree lies between the animals, fungi and amoebozoa (together called unikonts) on the one side, and plants, algae and most protozoa (bikonts) on the other. In fungi and animals, the genes for dihydrofolate reductase (DHFR) and thymidylate synthase (TS) are separate⁴⁴, as they are in prokaryote outgroups; but they are fused in the bikonts sampled so far. Assuming that the fusion occurred only once and that its subsequent fission did not occur at all, the DHFR–TS fusion would be a derived feature uniting bikonts, suggesting that the eukaryote root lies outside this group⁴⁹. The coherence of animals, fungi and various unicellular eukaryotes (together called opisthokonts) is supported by phylogenetic trees and other characters⁵⁰. The presence of a type II myosin in opisthokonts and amoebozoa unites them to form the unikonts⁵¹. If both unikonts and bikonts are monophyletic groups, and together they encompass extant eukaryotic diversity, then the root of eukaryotes would lie between them.

Placing the eukaryote root between unikonts and bikonts would help to bring order to chaos, if it is correct. However, it assumes that the underlying tree—over which the rooting character is mapped—is known, when in fact the relationships—especially for bikonts and many enigmatic protistan lineages⁵²—remain uncertain. The rooting also depends upon a single character of unknown stability sampled from only a few species. An additional caveat is that *Giardia* and *Trichomonas* lack both DHFR and TS—parasites relinquish genes of various biosynthetic pathways, stealing the pathway products from their hosts instead. Hence, the missing fusion character does not address their position in the tree.

New hypotheses of eukaryotic relationships

New data and analyses from many laboratories have been used to formulate a number of hypotheses of eukaryotic relationships (Fig. 3) that fundamentally differ from those in the SSU rRNA tree. It is apparent that hydrogenosomes and mitosomes appear on different branches; the absence of traditional mitochondria and presence of a specialized anaerobic phenotype are neither rare nor 'primitive', as once thought. Mitochondria with a genome encoding elements of the respiratory pathway also appear on both sides of the tree (Fig. 3), suggesting that this pathway has been retained since earliest times; although, as modern examples attest^{16,17}, it need not have always used oxygen as the sole terminal electron acceptor. On the basis of the unfolding tree, it would seem entirely possible—if not likely—that aerobic and anaerobic eukaryotes, harbouring mitochondrial homologues of various sorts, have co-existed throughout eukaryote history.

The relationships between major groups of eukaryotes are uncertain because of the lack of agreement between different proteins and different analyses; this uncertainty is depicted as a series of polytomies in Fig. 3. Most groups are still poorly sampled for species and molecular sequences—factors that impede robust resolution⁵³. It has been suggested⁵⁴ that the lack of resolution in deeper parts of the eukaryotic tree stems from an evolutionary 'big bang' or rapid radiation for eukaryotes, perhaps driven by the mitochondrial endosymbiosis⁵⁴. However, both theory and computer simulations^{40,41} suggest that a lack of resolution at deeper levels is to be expected given sparse data, our assumptions about sequence evolution, and the limitations of current phylogenetic methods. Thus, loss of historical signal provides a simple null hypothesis for the observed lack of resolution in deeper parts of the eukaryotic tree.

More good theories for eukaryotic origins than good data

Eukaryotic cell organization is more complex than prokaryotic, boasting, *inter alia*, a nucleus with its contiguous endoplasmic reticulum, Golgi, flagella with a '9+2' pattern of microtubule arrangement, and organelles surrounded by double membranes. There are no obvious precursor structures known among prokaryotes from which

such attributes could be derived, and no intermediate cell types known that would guide a gradual evolutionary inference between the prokaryotic and eukaryotic state. Accordingly, thoughts on the topic are diverse, and new suggestions appear faster than old ones can be tested.

Biologists have traditionally derived the complex eukaryotic state from the simpler prokaryotic one. In recent years, even that has been

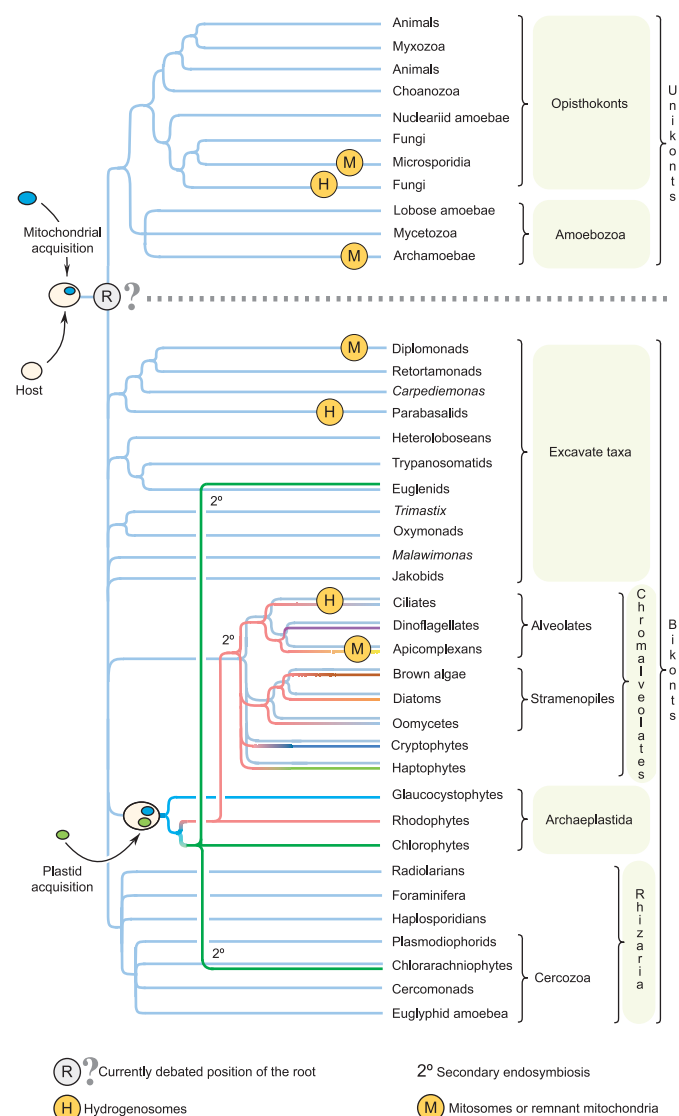


Figure 3 | Schematic tree of newer hypotheses for phylogenetic relationships among major groups of eukaryotes. The composite tree is based on work from many different laboratories and is summarized elsewhere⁵²; no single data set supports all branches. Polytomies indicate uncertainty in the branching order between major groups. The naming of groups follows current popular usage^{52,60}. The current debate that the root of the tree may split eukaryotes into bikonts and unikonts is discussed in the text. Lineages containing species with comparatively well-studied hydrogenosomes (H) or mitosomes (M) are labelled. The depicted distribution of hydrogenosomes and mitosomes is almost certainly conservative, as relatively few anaerobic or parasitic microbial eukaryotes have been studied in sufficient detail to characterize their organelles. The strict coevolution of host nuclear and algal nuclear plus plastid genomes within the confines of a single cell in the wake of secondary endosymbiosis (2°), irrespective of whether or not the secondary nucleus or plastid has persisted as a separate compartment, is indicated by doubled branches. Diversity of pigmentation among photosynthetic eukaryote lineages is symbolized by different coloured branches.

called into question, as some phylogenies have suggested that prokaryotes might be derived from eukaryotes⁵⁵. However, the ubiquity of mitochondrial homologues represents a strong argument that clearly polarizes the prokaryote-to-eukaryote transition: because the common ancestor of contemporary eukaryotes contained a mitochondrial endosymbiont that originated from within the proteobacterial lineage, we can confidently infer that prokaryotes arose and diversified before contemporary eukaryotes—the only ones whose origin requires explanation—did. This view is consistent with microfossil and biogeochemical evidence⁵⁶.

Current ideas on the origin of eukaryotes fall into two general classes: those that derive a nucleus-bearing but amitochondriate cell first, followed by the origin of mitochondria in a eukaryotic host^{57–61} (Fig. 4a–d), and those that derive the origin of mitochondria in a prokaryotic host, followed by the origin of eukaryotic-specific features^{62–64} (Fig. 4e–g). Models that derive a nucleated but amitochondriate cell as an intermediate (Fig. 4a–d) have suffered a substantial blow with the demise of Archezoa. Models that do not entail amitochondriate intermediates have in common that the host assumed to have acquired the mitochondrion was an archaeobacterium not a eukaryote; hence, the steep organizational grade between prokaryotes and eukaryotes follows in the wake of radical chimaer-

ism involving mitochondrial origins (Fig. 4e–g). A criticism facing all 'archaeobacterial host' models is that phagotrophy (the ability to engulf bacteria as food particles) was once seen as an absolute prerequisite for mitochondrial origins⁶⁰. This argument has lost some of its strength with the discovery of symbioses where one prokaryote lives inside another, non-phagocytotic prokaryote⁶⁵.

The elusive informational ancestor

With the exception of the neomuran hypothesis, which views both eukaryotes and archaeobacteria as descendants of Gram-positive eubacteria^{60,61} (Fig. 4d), most current theories for eukaryotic origins (Fig. 4) posit the involvement of an archaeobacterium in that process. The archaeobacterial link to eukaryote origins was first inferred from shared immunological and biochemical similarities of their DNA-dependent RNA polymerases⁶⁶. Tree-based studies of entire genomes^{67,68} extended this observation: most eukaryotic genes for replication, transcription and translation (informational genes) are related to archaeobacterial homologues, while those encoding biosynthetic and metabolism functions (operational genes) are usually related to eubacterial homologues^{8,67,68}.

The rooted SSU rRNA tree¹ depicts eukaryotes and archaeobacteria as sister groups, as in the neomuran (Fig. 4d) hypothesis^{60,61}. By

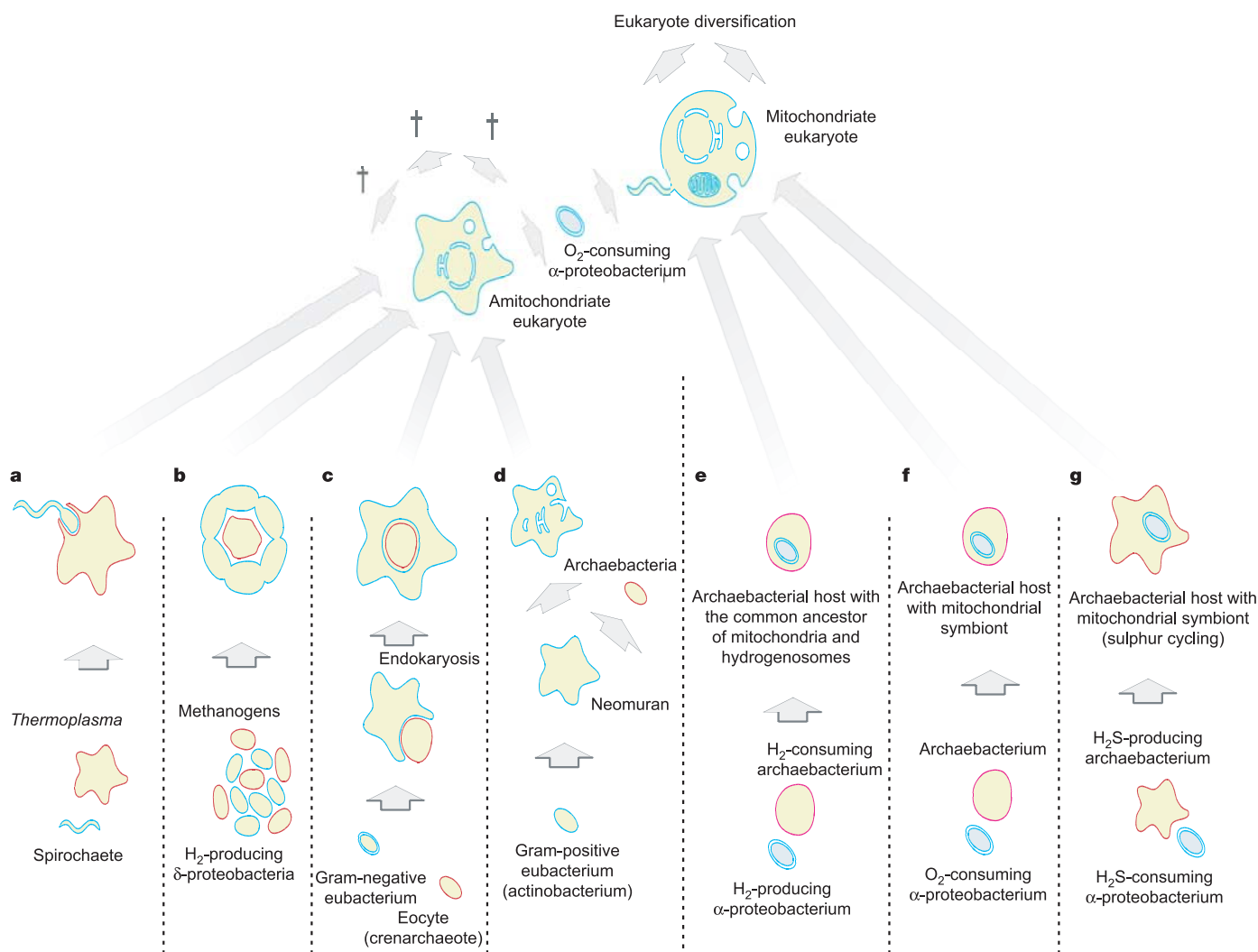


Figure 4 | Models for eukaryote origins that are, in principle, testable with genome data. **a–d**, Models that propose the origin of a nucleus-bearing but amitochondriate cell first, followed by the acquisition of mitochondria in a eukaryotic host. **e–g**, Models that propose the origin of mitochondria in a prokaryotic host, followed by the acquisition of eukaryotic-specific

features. Panels **a–g** are redrawn from refs 57 (**a**), 58 (**b**), 59 (**c**), 60 and 61 (**d**), 62 (**e**), 63 (**f**) and 64 (**g**). The relevant microbial players in each model are labelled. Archaeobacterial and eubacterial lipid membranes are indicated in red and blue, respectively.

contrast, the eocyte (Fig. 4c) hypothesis^{69,70} proposes that eukaryotic informational genes originate from a specific lineage of archaeobacteria called the eocytes, a group synonymous with the Crenarchaeota¹. In the eocyte tree, the eukaryotic genetic machinery is descended from within the archaeobacteria. Although the rooted rRNA tree is vastly more visible to non-specialists, published data are equivocal: for every analysis of a eukaryotic informational gene that recovers the neomuran topology, a different analysis of the same molecule(s) has recovered the eocyte tree^{70–74}, with the latter being favoured by more sophisticated phylogenetic analyses^{69,73,74} and by a shared amino-acid insertion in eocyte and eukaryotic elongation factor 1- α ⁷⁰.

More recently, genome trees based on shared gene content have been reported. These methods are still new, and—just like gene trees—give different answers from the same data, recovering for informational genes either eukaryote–archaeobacterial sisterhood⁷⁵, the eocyte tree⁷⁶ or a euryarchaeote ancestry⁷⁷. The dichotomy of archaeobacteria into euryarchaeotes and eocytes/crenarchaeotes¹ remains unchallenged. The issue, so far unresolved, is the relationship of eukaryotic informational genes to archaeobacterial homologues: inheritance from a common progenitor (as in the neomuran hypothesis) or a direct descendant; and if by direct descent, from eocytes/crenarchaeotes like *Sulfolobus*⁷⁶, or euryarchaeotes such as *Thermoplasma*^{64,78}, *Pyrococcus*⁷⁷ or methanogens^{58,62}. The problems associated with the phylogenetic relationships discussed above are exacerbated at such deep levels, and there is currently neither consensus on this issue nor unambiguous evidence that would clarify it.

The vexing operational majority

Of those eukaryotic genes that have detectable prokaryotic homologues, the majority⁶⁷, perhaps as much as 75%⁸, are eubacterial and correspond to the operational class. Here arises an interesting point. Although individual analyses of informational genes arrive at fundamentally different interpretations^{76,77}, no one has yet suggested that more than one archaeobacterium participated in eukaryote origins. The situation is quite different with operational genes, where differing phylogenies for individual genes are freely interpreted as evidence for the participation of more than one eubacterial partner. The contribution of gene transfers from the ancestral mitochondrion to nuclear chromosomes has been estimated as anywhere from 136–157 (ref. 77) to ~630 genes⁷⁹, depending on the method of analysis. An issue that still requires clarification concerns the origin of thousands of eukaryotic operational genes that are clearly eubacterial, but not specifically α -proteobacterial, in origin⁸ (disregarding here the cyanobacterial genes in plants⁸⁰).

There are currently four main theories that attempt to account for those genes. (1) In the neomuran hypothesis (Fig. 4d), they are explained through a direct inheritance from the Gram-positive ancestor^{60,61}; however, few eukaryote genes branch with Gram-positive homologues. (2) In hypotheses entailing more than one eubacterial partner at eukaryote origins (Fig. 4a–c), they are explained as descending from the non-mitochondrial eubacterium; however, these genes branch all over the eubacterial tree, not with any particular lineage. (3) In models favouring widespread LGT from prokaryotes to eukaryotes, they are explained as separate acquisitions from individual donors⁸¹; although some LGT clearly has occurred⁸², the jury is still out on its extent because of a lack of detailed large-scale analyses of individual genes using reliable methods. (4) In single-eubacterium models (Fig. 4e–g), they are either not addressed, or explained as acquisitions from the mitochondrial symbiont, with a twofold corollary⁸ of LGT among free-living prokaryotes since the origin of mitochondria, and phylogenetic artefact.

LGT among prokaryotes⁸³ figures into the origin of eukaryotic operational genes in a fundamental manner that is often overlooked. Most claims of outright LGT to ancestral eukaryotes (that is, from donors distinct from the mitochondrion) implicitly assume a static

chromosome model in which prokaryotes do not exchange genes among themselves; finding a eukaryotic gene that branches with a group other than α -proteobacteria is taken as evidence for an origin from that group (the vagaries of deep branches notwithstanding). But if we embrace a fluid chromosome model for prokaryotes, as some interpretations of the data suggest we should⁸⁴, then the expected phylogeny for a gene acquired from the mitochondrion would be common ancestry for all eukaryotes, but not necessarily tracing to α -proteobacteria, because the ancestor of mitochondria possessed an as yet unknown collection of genes.

The timing and ecological context of eukaryote origins

Diversified unicellular microfossils of uncertain phylogenetic affinity (acritarchs), but widely accepted as eukaryotes, appear in strata of ~1.45 billion years (Gyr) of age⁸⁵, providing a minimum age for the group. *Bangiomorpha*, a fossilized multicellular organism virtually indistinguishable in morphology from modern bangiophyte red algae, has been found in strata of ~1.2 Gyr of age⁸⁶, placing a lower bound on the age of the plant kingdom. A wide range of molecular clock estimates of eukaryote age have been reported, but these are still uncertain, being contingent both on the use of younger calibration points and on the phylogenetic model and assumed tree⁸⁷. At present, a minimum age of eukaryotes at ~1.45 Gyr and a minimum age of the plant kingdom at ~1.2 Gyr seem to be criteria that the molecular clock must meet.

The classical view of early eukaryote evolution posits two main ecological stages: (1) the early emergence and diversification of anaerobic, amitochondriate lineages, followed by (2) the acquisition of an oxygen-respiring mitochondrial ancestor in one lineage thereof and the subsequent diversification of aerobic eukaryotic lineages⁷⁸. Concordant with that view, mitochondrial origins have traditionally been causally linked to the global rise in atmospheric oxygen levels at ~2 Gyr ago and an assumed 'environmental disaster' for cells lacking the mitochondrial endosymbiont^{63,88}, providing a selective force (oxygen detoxification) for the acquisition of the mitochondrion^{63,88}. Two observations challenge this model.

First, it is now clear that the contemporary anaerobic eukaryotes did not branch off before the origin of mitochondria. Second, new isotope studies indicate that anaerobic environments persisted locally and globally over the past 2 Gyr. That oxygen first appeared in the atmosphere at ~2 Gyr ago is still generally accepted, but it is now thought that, up until about 600 Myr ago, the oceans existed in an intermediate oxidation state, with oxygenated surface water (where photosynthesis was occurring), and sulphide-rich (sulphidic) and oxygen-lacking (anoxic) subsurface water^{89,90}. Hence, the 'oxygen event' in the atmosphere should be logically decoupled from anoxic marine environments, where anaerobic eukaryotes living on the margins of an oxic world could have flourished, as they still do today²⁷.

Outlook

In the past, phylogenetic trees have produced a particular view of early eukaryote history that was appealing, but turned out to be wrong in salient aspects. Simply testing whether a model used to make a tree actually fits the data⁴⁰ would do much to restore confidence in the merits of deep phylogenetic analyses. The fact that monophyly of plants can be recovered using molecular sequences⁹¹, an event that should predate 1.2 Gyr, suggests that ancient signal can be extracted, but how far back we might expect to be able to go is uncertain. The persistence of mitochondrially derived organelles in all eukaryotes, and plastids in some lineages, provides phylogeny-independent evidence for the occurrence of those symbiotic events. But independent evidence for the participation of other prokaryotic endosymbionts is lacking. Analysis of mitochondria in their various guises has revealed that their unifying trait is neither respiration nor ATP synthesis; the common essential function, if any, for contemporary eukaryotes remains to be pinpointed by

comparative study. It may still be that a eukaryote is lurking out there that never possessed a mitochondrion—a *bona fide* archezoan—in which case prokaryote-host models (Fig. 4e–g) for eukaryogenesis can be abandoned. However, morphological studies and environmental sequencing efforts performed so far from the best candidate habitats to harbour such relics—anaerobic marine sediments—have not uncovered new, unknown and more-deeply branching lineages; rather, they have uncovered a greater diversity of lineages with affinities to known mitochondriate groups^{28,61}. The available phylogenetic findings from genomes are not fully consistent with any current hypothesis for eukaryote origins, the underlying reasons for which—biological, methodological or both—are as yet unclear. Genomes must surely bear some testimony to eukaryotic origins, but new approaches and more rigorous attention to the details of phylogenetic inference will be required to decipher the message.

- Woese, C. R., Kandler, O. & Wheelis, M. L. Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl Acad. Sci. USA* **87**, 4576–4579 (1990).
- Leipe, D. D., Gunderson, J. H., Nerad, T. A. & Sogin, M. L. Small subunit ribosomal RNA of *Hexamita inflata* and the quest for the first branch in the eukaryotic tree. *Mol. Biochem. Parasitol.* **59**, 41–48 (1993).
- Hashimoto, T., Nakamura, Y., Kamaishi, T. & Hasegawa, M. Early evolution of eukaryotes inferred from the amino acid sequences of elongation factors 1 α and 2. *Arch. Protistenkd.* **148**, 287–295 (1997).
- Cavalier-Smith, T. in *Endocytobiology II* (eds Schwemmler, W. & Schenk, H. E. A.) 1027–1034 (De Gruyter, Berlin, 1983).
- Gray, M. W., Lang, B. F. & Burger, G. Mitochondria of protists. *Annu. Rev. Genet.* **38**, 477–524 (2004).
- Cavalier-Smith, T. in *Endocytobiology II* (eds Schwemmler, W. & Schenk, H. E. A.) 265–279 (De Gruyter, Berlin, 1983).
- Planner, N. & Geissler, A. Versatility of the mitochondrial protein import machinery. *Nature Rev. Mol. Cell Biol.* **2**, 339–349 (2001).
- Esser, C. et al. A genome phylogeny for mitochondria among α -proteobacteria and a predominantly eubacterial ancestry of yeast nuclear genes. *Mol. Biol. Evol.* **21**, 1643–1660 (2004).
- Timmis, J. N., Ayliffe, M. A., Huang, C. Y. & Martin, W. Endosymbiotic gene transfer: Organelle genomes forge eukaryotic chromosomes. *Nature Rev. Genet.* **5**, 123–135 (2004).
- Clark, C. G. & Roger, A. J. Direct evidence for secondary loss of mitochondria in *Entamoeba histolytica*. *Proc. Natl Acad. Sci. USA* **92**, 6518–6521 (1995).
- Roger, A. J. Reconstructing early events in eukaryotic evolution. *Am. Nat.* **154**, S146–S163 (1999).
- Abrahamson, M. S. et al. Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. *Science* **304**, 441–445 (2004).
- Boxma, B. et al. An anaerobic mitochondrion that produces hydrogen. *Nature* **434**, 74–79 (2005).
- Müller, M. in *Molecular Medical Parasitology* (eds Marr, J. J., Nilsen, T. W. & Komuniecki, R. W.) 125–139 (Academic, Amsterdam, 2003).
- Katinka, M. D. et al. Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. *Nature* **414**, 450–453 (2001).
- Tielens, A. G., Rotte, C., van Hellemond, J. J. & Martin, W. Mitochondria as we don't know them. *Trends Biochem. Sci.* **27**, 564–572 (2002).
- Komuniecki, R. W. & Tielens, A. G. M. in *Molecular Medical Parasitology* (eds Marr, J. J., Nilsen, T. W. & Komuniecki, R.) 339–358 (Academic, Amsterdam, 2003).
- Darwin, C. *The Origin of Species* Reprint edn (Penguin Books, London, 1968).
- Müller, M. The hydrogenosome. *J. Gen. Microbiol.* **139**, 2879–2889 (1993).
- van der Giezen, M. et al. Conserved properties of hydrogenosomal and mitochondrial ADP/ATP carriers: A common origin for both organelles. *EMBO J.* **21**, 572–579 (2002).
- Tjaden, J. et al. A divergent ADP/ATP carrier in the hydrogenosomes of *Trichomonas gallinae* argues for an independent origin of these organelles. *Mol. Microbiol.* **51**, 1439–1446 (2004).
- Embley, T. M. et al. Hydrogenosomes, mitochondria and early eukaryotic evolution. *IUBMB Life* **55**, 387–395 (2003).
- Dyall, S. D. et al. Presence of a member of the mitochondrial carrier family in hydrogenosomes: Conservation of membrane-targeting pathways between hydrogenosomes and mitochondria. *Mol. Cell. Biol.* **20**, 2488–2497 (2000).
- Sutak, R. et al. Mitochondrial-type assembly of FeS centers in the hydrogenosomes of the amitochondriate eukaryote *Trichomonas vaginalis*. *Proc. Natl Acad. Sci. USA* **101**, 10368–10373 (2004).
- Hrdy, I. et al. *Trichomonas* hydrogenosomes contain the NADH dehydrogenase module of mitochondrial complex I. *Nature* **432**, 618–622 (2004).
- Schnarrenberger, C. & Martin, W. Evolution of the enzymes of the citric acid cycle and the glyoxylate cycle of higher plants. A case study of endosymbiotic gene transfer. *Eur. J. Biochem.* **269**, 868–883 (2002).
- Fenchel, T. & Finlay, B. J. *Ecology and Evolution in Anoxic Worlds* (eds May, R. M. & Harvey, P. H.) (Oxford Univ. Press, Oxford, 1995).
- Roger, A. J. & Silberman, J. D. Cell evolution: Mitochondria in hiding. *Nature* **418**, 827–829 (2002).
- Whatley, J. M., John, P. & Whatley, F. R. From extracellular to intracellular: The establishment of mitochondria and chloroplasts. *Proc. R. Soc. Lond. B* **204**, 165–187 (1979).
- Dyall, S. D., Brown, M. T. & Johnson, P. J. Ancient invasions: From endosymbionts to organelles. *Science* **304**, 253–257 (2004).
- Dyall, S. D. et al. Non-mitochondrial complex I proteins in a hydrogenosomal oxidoreductase complex. *Nature* **431**, 1103–1107 (2004).
- Tovar, J., Fischer, A. & Clark, C. G. The mitosome, a novel organelle related to mitochondria in the amitochondrial parasite *Entamoeba histolytica*. *Mol. Microbiol.* **32**, 1013–1021 (1999).
- Tovar, J. et al. Mitochondrial remnant organelles of *Giardia* function in iron-sulphur protein maturation. *Nature* **426**, 172–176 (2003).
- Williams, B. A., Hirt, R. P., Lucocq, J. M. & Embley, T. M. A mitochondrial remnant in the microsporidian *Trachipleistophora hominis*. *Nature* **418**, 865–869 (2002).
- Regoes, A. et al. Protein import, replication and inheritance of a vestigial mitochondrion. *J. Biol. Chem.* **280**, 30557–30563 (2005).
- Chan, K. W. et al. A Novel ADP/ATP transporter in the mitosome of the microaerophilic human parasite *Entamoeba histolytica*. *Curr. Biol.* **15**, 737–742 (2005).
- Dolezal, P. et al. *Giardia* mitosomes and trichomonad hydrogenosomes share a common mode of protein targeting. *Proc. Natl Acad. Sci. USA* **102**, 10924–10929 (2005).
- Lill, R. & Muhlenhoff, U. Iron-sulfur-protein biogenesis in eukaryotes. *Trends Biochem. Sci.* **30**, 133–141 (2005).
- Ali, V., Shigeta, Y., Tokumoto, U., Takahashi, Y. & Nozaki, T. An intestinal parasitic protist, *Entamoeba histolytica*, possesses a non-redundant nitrogen fixation-like system for iron-sulfur cluster assembly under anaerobic conditions. *J. Biol. Chem.* **279**, 16863–16874 (2004).
- Penny, D., McComish, B. J., Charleston, M. A. & Hendy, M. D. Mathematical elegance with biochemical realism: The covarian model of molecular evolution. *J. Mol. Evol.* **53**, 711–723 (2001).
- Ho, S. Y. W. & Jermini, L. S. Tracing the decay of the historical signal in biological sequence data. *Syst. Biol.* **53**, 623–637 (2004).
- Felsenstein, J. Cases in which parsimony or incompatibility methods will be positively misleading. *Syst. Zool.* **25**, 401–410 (1978).
- Stiller, J. W. & Hall, B. D. Long-branch attraction and the rDNA model of early eukaryotic evolution. *Mol. Biol. Evol.* **16**, 1270–1279 (1999).
- Philippe, H. et al. Early-branching or fast-evolving eukaryotes? An answer based on slowly evolving positions. *Proc. R. Soc. Lond. B* **267**, 1213–1221 (2000).
- Hirt, R. P. et al. Microsporidia are related to fungi: Evidence from the largest subunit of RNA polymerase II and other proteins. *Proc. Natl Acad. Sci. USA* **96**, 580–585 (1999).
- Keeling, P. J., Luker, M. A. & Palmer, J. D. Evidence from beta-tubulin phylogeny that microsporidia evolved from within the fungi. *Mol. Biol. Evol.* **17**, 23–31 (2000).
- Arisue, N., Hasegawa, M. & Hashimoto, T. Root of the Eukaryota tree as inferred from combined maximum likelihood analyses of multiple molecular sequence data. *Mol. Biol. Evol.* **22**, 409–420 (2005).
- Penny, D. Criteria for optimising phylogenetic trees and the problem of determining the root of a tree. *J. Mol. Evol.* **8**, 95–116 (1976).
- Stechmann, A. & Cavalier-Smith, T. The root of the eukaryote tree pinpointed. *Curr. Biol.* **13**, R665–R666 (2003).
- Steenkamp, E. T., Wright, J. & Baldauf, S. L. The protistan origins of animals and fungi. *Mol. Biol. Evol.* **23**, 93–106 (2006); published online 8 September 2005 (doi:10.1093/molbev/msj011).
- Richards, T. A. & Cavalier-Smith, T. Myosin domain evolution and the primary divergence of eukaryotes. *Nature* **436**, 1113–1118 (2005).
- Adl, S. M. et al. The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. *J. Eukaryot. Microbiol.* **52**, 399–451 (2005).
- Graybeal, A. Is it better to add taxa or characters to a difficult phylogenetic problem? *Syst. Biol.* **47**, 9–17 (1998).
- Philippe, H. & Adoutte, A. in *Evolutionary Relationships Among Protozoa* (eds Coombs, G. H., Vickerman, K., Sleigh, M. A. & Warren, A.) 25–56 (Kluwer Academic, Dordrecht, 1998).
- Forterre, P. & Philippe, H. Where is the root of the universal tree of life? *Bioessays* **21**, 871–879 (1999).
- Knoll, A. H. *Life on a Young Planet: The First Three Billion Years of Evolution on Earth* (Princeton Univ. Press, 2003).
- Margulis, L., Dolan, M. F. & Whiteside, J. H. "Imperfections and oddities" in the origin of the nucleus. *Paleobiology* **31**, 175–191 (2005).
- Moreira, D. & Lopez Garcia, P. Symbiosis between methanogenic archaea and δ -proteobacteria as the origin of eukaryotes: The syntrophic hypothesis. *J. Mol. Evol.* **47**, 517–530 (1998).
- Lake, J., Moore, J., Simonson, A. & Rivera, M. in *Microbial Phylogeny and Evolution Concepts and Controversies* (ed. Sapp, J.) 184–206 (Oxford Univ. Press, Oxford, 2005).

60. Cavalier-Smith, T. The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. *Int. J. Syst. Evol. Microbiol.* **52**, 297–354 (2002).
61. Cavalier-Smith, T. Only six kingdoms of life. *Proc. R. Soc. Lond. B* **271**, 1251–1262 (2004).
62. Martin, W. & Müller, M. The hydrogen hypothesis for the first eukaryote. *Nature* **392**, 37–41 (1998).
63. Vellai, T., Takacs, K. & Vida, G. A new aspect to the origin and evolution of eukaryotes. *J. Mol. Evol.* **46**, 499–507 (1998).
64. Searcy, D. G. in *The Origin and Evolution of the Cell* (eds Matsuno, H. H. & Matsuno, K.) 47–78 (World Scientific, Singapore, 1992).
65. von Dohlen, C. D., Kohler, S., Alsop, S. T. & McManus, W. R. Mealybug β -proteobacterial endosymbionts contain γ -proteobacterial symbionts. *Nature* **412**, 433–436 (2001).
66. Zillig, W., Schnabel, R. & Stetter, K. O. Archaeobacteria and the origin of the eukaryotic cytoplasm. *Curr. Top. Microbiol. Immunol.* **114**, 1–18 (1985).
67. Rivera, M. C., Jain, R., Moore, J. E. & Lake, J. A. Genomic evidence for two functionally distinct gene classes. *Proc. Natl Acad. Sci. USA* **95**, 6239–6244 (1998).
68. Ribeiro, S. & Golding, G. B. The mosaic nature of the eukaryotic nucleus. *Mol. Biol. Evol.* **15**, 779–788 (1998).
69. Lake, J. A. Origin of the eukaryotic nucleus determined by rate-invariant analysis of rRNA sequences. *Nature* **331**, 184–186 (1988).
70. Rivera, M. C. & Lake, J. A. Evidence that eukaryotes and eocyte prokaryotes are immediate relatives. *Science* **257**, 74–76 (1992).
71. Baldauf, S. L., Palmer, J. D. & Doolittle, W. F. The root of the universal tree and the origin of eukaryotes based upon elongation factor phylogeny. *Proc. Natl Acad. Sci. USA* **93**, 7749–7754 (1996).
72. Brown, J. R. & Doolittle, W. F. Archaea and the prokaryote-to-eukaryote transition. *Microbiol. Mol. Biol. Rev.* **61**, 456–502 (1997).
73. Tourasse, N. J. & Gouy, M. Accounting for evolutionary rate variation among sequence sites consistently changes universal phylogenies deduced from rRNA and protein-coding genes. *Mol. Phylogenet. Evol.* **13**, 159–168 (1999).
74. Brown, J. R. *et al.* Universal trees based on large combined protein sequence data sets. *Nature Genet.* **28**, 281–285 (2001).
75. Daubin, V., Gouy, M. & Perriere, G. A phylogenomic approach to bacterial phylogeny: Evidence of a core of genes sharing a common history. *Genome Res.* **12**, 1080–1090 (2002).
76. Rivera, M. C. & Lake, J. A. The ring of life provides evidence for a genome fusion origin of eukaryotes. *Nature* **431**, 152–155 (2004).
77. Horiike, T., Hamada, K., Miyata, D. & Shinozawa, T. The origin of eukaryotes is suggested as the symbiosis of *Pyrococcus* into γ -proteobacteria by phylogenetic tree based on gene content. *J. Mol. Evol.* **59**, 606–619 (2004).
78. Margulis, L., Dolan, M. F. & Guerrero, R. The chimeric eukaryote: Origin of the nucleus from the karyomastigont in amitochondriate protists. *Proc. Natl Acad. Sci. USA* **97**, 6954–6959 (2000).
79. Gabaldon, T. & Huynen, M. A. Reconstruction of the proto-mitochondrial metabolism. *Science* **301**, 609 (2003).
80. Martin, W. *et al.* Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc. Natl Acad. Sci. USA* **99**, 12246–12251 (2002).
81. Doolittle, W. F. You are what you eat: A gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. *Trends Genet.* **14**, 307–311 (1998).
82. Loftus, B. *et al.* The genome of the protist parasite *Entamoeba histolytica*. *Nature* **433**, 865–868 (2005).
83. Doolittle, W. F. Lateral genomics. *Trends Cell Biol.* **9**, M5–M8 (1999).
84. Kunin, V., Goldovsky, L., Darzentas, N. & Ouzounis, C. A. The net of life—reconstruction of the microbial phylogenetic network. *Genome Res.* **15**, 954–959 (2005).
85. Javaux, E. J., Knoll, A. H. & Walter, M. R. Morphological and ecological complexity in early eukaryotic ecosystems. *Nature* **412**, 66–69 (2001).
86. Butterfield, N. J. *Bangiomorpha pubescens* n. gen., n. sp.: implications for the evolution of sex, multicellularity, and the Mesoproterozoic/Neoproterozoic radiation of eukaryotes. *Paleobiology* **26**, 386–404 (2000).
87. Benton, M. J. & Ayala, F. J. Dating the tree of life. *Science* **300**, 1698–1700 (2003).
88. Kurland, C. G. & Andersson, S. G. Origin and evolution of the mitochondrial proteome. *Microbiol. Mol. Biol. Rev.* **64**, 786–820 (2000).
89. Shen, Y., Knoll, A. H. & Walter, M. R. Evidence for low sulphate and anoxia in a mid-Proterozoic marine basin. *Nature* **423**, 632–635 (2003).
90. Poulton, S. W., Fralick, P. W. & Canfield, D. E. The transition to a sulphidic ocean ~1.84 billion years ago. *Nature* **431**, 173–177 (2004).
91. Rodríguez-Ezpeleta, N. *et al.* Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. *Curr. Biol.* **15**, 1325–1330 (2005).
92. Pace, N. R. A molecular view of microbial diversity and the biosphere. *Science* **276**, 734–740 (1997).
93. Keithly, J. S., Langreth, S. G., Buttle, K. F. & Mannella, C. A. Electron tomographic and ultrastructural analysis of the *Cryptosporidium parvum* relict mitochondrion, its associated membranes, and organelles. *J. Eukaryot. Microbiol.* **52**, 132–140 (2005).

Acknowledgements We thank M. Müller, J. Archibald, R. Hirt, K. Henze and L. Tielens, and members of our laboratories, for discussions.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence should be addressed to T.M.E. (martin.embley@ncl.ac.uk) or W.M. (w.martin@uni-duesseldorf.de).

Proteome survey reveals modularity of the yeast cell machinery

Anne-Claude Gavin^{1*†}, Patrick Aloy^{2*}, Paola Grandi¹, Roland Krause^{1,3}, Markus Boesche¹, Martina Marzioch¹, Christina Rau¹, Lars Juhl Jensen², Sonja Bastuck¹, Birgit Dimpelfeld¹, Angela Edelmann¹, Marie-Anne Heurtier¹, Verena Hoffman¹, Christian Hoefert¹, Karin Klein¹, Manuela Hudak¹, Anne-Marie Michon¹, Malgorzata Schelder¹, Markus Schirle¹, Marita Remor¹, Tatjana Rudi¹, Sean Hooper², Andreas Bauer¹, Tewis Bouwmeester¹, Georg Casari¹, Gerard Drewes¹, Gitte Neubauer¹, Jens M. Rick¹, Bernhard Kuster¹, Peer Bork², Robert B. Russell² & Giulio Superti-Furga^{1,4}

Protein complexes are key molecular entities that integrate multiple gene products to perform cellular functions. Here we report the first genome-wide screen for complexes in an organism, budding yeast, using affinity purification and mass spectrometry. Through systematic tagging of open reading frames (ORFs), the majority of complexes were purified several times, suggesting screen saturation. The richness of the data set enabled a *de novo* characterization of the composition and organization of the cellular machinery. The ensemble of cellular proteins partitions into 491 complexes, of which 257 are novel, that differentially combine with additional attachment proteins or protein modules to enable a diversification of potential functions. Support for this modular organization of the proteome comes from integration with available data on expression, localization, function, evolutionary conservation, protein structure and binary interactions. This study provides the largest collection of physically determined eukaryotic cellular machines so far and a platform for biological data integration and modelling.

Genomes are remarkable in that they encode most of the functions necessary for their interpretation and propagation¹. However, many principles as to how individual gene products form the structures required for biological activity are still unknown. Biological processes, such as the cell cycle and replication, require precise organization of molecules in time and space. Complexes are among the fundamental units of macromolecular organization². They are thought to assemble in a particular order, and often require energy-driven conformational changes, specific post-translational modifications or chaperone assistance for proper formation³. Their composition is also known to vary according to cellular requirements.

Affinity purification methods are well suited for studying complexes under near-physiological conditions^{4,5}. They allow macromolecules physically associated with a tagged bait to be retrieved and identified by mass spectrometry^{6,7}. These methods have been applied as large-scale screens in prokaryotic and eukaryotic cells, and have led to a growing collection of cellular machines^{8–11} that, in combination with large-scale yeast two-hybrid studies^{12,13}, are powerful integrators of additional biological data^{14–16}. However, in the absence of a genome-wide screen, where many complexes are retrieved repeatedly through a 'reverse purification' process, assignment of a component to a particular complex relied heavily on experimental stringency and arbitrary thresholds. Here we report the first genome-wide screen for complexes to investigate the underlying organizational principles of the eukaryotic cellular machinery.

Genome-wide characterization of complexes

We applied the tandem-affinity-purification method coupled to

mass spectrometry (TAP-MS)^{6–8} to all 6,466 ORFs of *Saccharomyces cerevisiae* as annotated in 2002 (refs 17, 18; Fig. 1 and Supplementary Information). We employed standardized protocols and successfully purified 1,993 unique TAP-fusion proteins, of which 88% retrieved at least one partner (Fig. 1; Supplementary Table S1). From all purifications, we processed 52,000 samples for mass spectrometry and identified 36,000 proteins, of which 2,760 were distinct (Fig. 1; Supplementary Figs S2–S5). These represent about 60% of the estimated proteome for exponentially growing yeast^{19–21}, and cover all functional classes and subcellular localizations. The absolute abundances of the identified proteins show a wide range, from 32 to 500,000 copies per cell¹⁹, although coverage varied considerably, being highest for the most abundant proteins (>16,000 copies per cell: 80% coverage), and lowest for the rarest proteins (<500 copies: 40% coverage) (Supplementary Fig. S1). We measured reproducibility by performing 139 purifications in duplicate (99 soluble; 40 membrane), and found that, on average, 69% of recovered proteins were common to both, giving an approximation of false-positive/negative rates within the raw data. However, as complexes are retrieved in several purifications, interactions observed repeatedly are more likely to be correct (see below).

The purification data contains 73% of known complexes from the Munich Information Center for Protein Sequences (MIPS) database²² (217 complexes) and our own literature mining (62 complexes). We found no evidence for 74 known complexes, possibly because they may not assemble under our growth conditions or because the tag interferes with complex assembly⁸. This is the case for the partially recovered CCT (chaperonin-containing tailless complex

¹Cellzome AG, Meyerhofstrasse 1, 69117 Heidelberg, Germany. ²EMBL, Meyerhofstrasse 1, 69117 Heidelberg, Germany. ³MPI-MG, MPI-IB, Charité Campus Mitte, Schumannstrasse 21/22, 10117 Berlin, Germany. ⁴Center for Molecular Medicine of the Austrian Academy of Sciences, Lazarettgasse 19, 1090 Vienna, Austria. [†]Present address: EMBL, Meyerhofstrasse 1, 69117 Heidelberg, Germany.

*These authors contributed equally to this work.

polypeptide 1) complex—the carboxy termini of the eight subunits in the ring-like core of the complex lie on interaction interfaces²³. However, these situations could often be rescued: 30% of TAP-tagged proteins that we could not purify were detected in purifications using other complex components.

We used a modified purification procedure for membrane proteins and successfully purified 340 of the 628 that were tagged. For example, we retrieved the Q/t-SNARE complex, including both integral membrane components of the trimeric receptor (Ues1, Sec20 and Ufe1) and the peripheral membrane machinery (Dsl1, Sec39, Tip20) required for stability²⁴. We also detected novel links such as that between the Akr1 palmitoyl transferase (a six-transmembrane-segment protein) and Ste4 (the G β subunit of the pheromone receptor-coupled G protein), which is consistent with genetic evidence²⁵ and supports a role for protein acylation in the pheromone response.

De novo definition of protein complexes

The proportion of new proteins identified per purification dropped asymptotically during the progression of the screen, suggesting that the procedure was to near saturation (Supplementary Fig. S6a). We also observed that 64% of known complexes²² were retrieved several times resulting in a high coverage of known components (Supplementary Fig. S6b). We exploited this redundancy to define complexes computationally. Current approaches for defining complexes from binary interactions²⁶ were not deemed appropriate as these are not directly inferable from purifications. We also explicitly avoided the incorporation of prior knowledge to circumvent any bias towards well-studied proteins.

We first derived a 'socio-affinity' index (see the Methods) that quantifies the propensity of proteins to form partnerships. It measures the log-odds of the number of times two proteins are observed together, relative to what would be expected from their frequency in the data set, and encompasses both the 'spoke' and the 'matrix' models for assigning binary interactions within purifications. The index accounts for the frequency of proteins within

the data set and thus naturally discriminates true from spurious interactions involving very promiscuous partners. For instance, Vma2, which was seen in 552 purifications and would have been ignored under previous high-frequency filtering strategies^{8,9}, showed high indices only with proteins it is known to associate with (Vma5, Vma6, Vma10 and Rav1). Generally, pairs with socio-affinity indices below 5 should be considered with caution (reproducibility <70%), though those above 5 are more reliable (89%). These indices capture some biochemical properties of protein–protein interactions: there is a tentative correlation with the few dissociation constants available in the literature ($P < 0.08$) and protein pairs with high socio-affinity indices are more likely to be in direct contact as measured either by three-dimensional structures or the yeast two-hybrid system (Supplementary Fig. S7). To our knowledge, this is the first attempt to re-create numbers approximating physical measurements purely from proteomics data.

If each protein only belonged to a single complex, we could generate a definitive set by a single clustering step using socio-affinity indices. However, it is well established that proteins can be present in multiple complexes; a property we reasoned could be captured by an iterative procedure. Briefly, we first used the socio-affinity indices to form a matrix for all pairs of proteins studied, and then applied cluster analysis to generate an initial list of complexes. We then subtracted a penalty from the initial matrix values and repeated clustering. Tight associations are not drastically affected by the penalty, while looser ones are gradually eroded, and can be replaced by others not present initially. We varied the clustering parameters (number of iterations, clustering type, penalty values, and so on) over a sensible range to produce 1,784 different complex sets, and compared each to a manually curated group of known complexes used for structural analysis¹⁴. We computed both coverage (that is, the fraction of proteins in known complexes that we retrieved) and accuracy (that is, the fraction of the retrieved complexes components that match those already known; Fig. 1). The best conditions generated a collection of 491 complexes with 83% coverage and 78% accuracy. However, inspection revealed that known complex

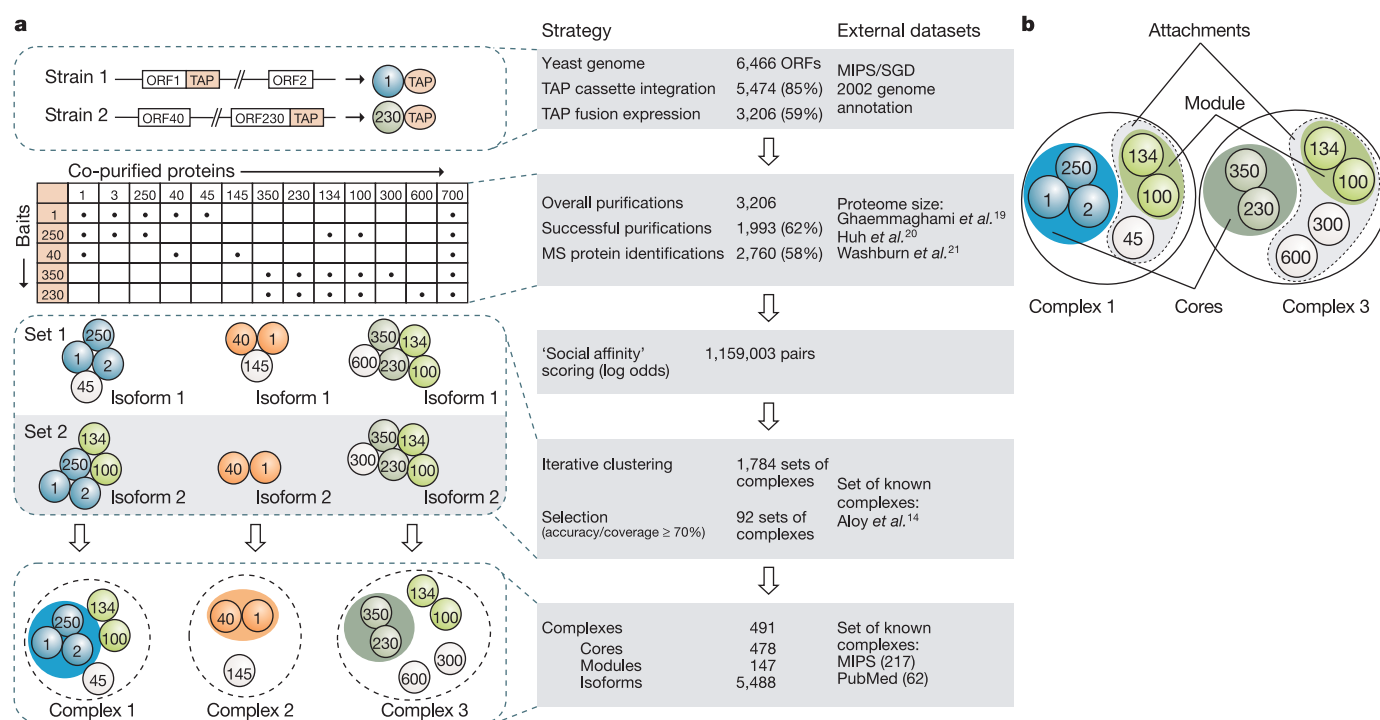


Figure 1 | Synopsis of the genome-wide screen for complexes and data analysis. a, Summary of the overall experimental strategy. MIPS/SGD, Munich Information Center for Protein Sequences/*Saccharomyces* Genome

Database. **b**, Definition and terminology used to define protein-complex architecture.

components could be found under clustering conditions with slightly poorer accuracy or coverage. Therefore, we grouped similar complexes from conditions with coverage and accuracy above 70%. The resulting 5,488 different protein-complex variations were termed ‘complex isoforms’ (Fig. 1). This procedure increased the overall coverage to 90%. The inclusion of parameters resulting in accuracy/coverage below 70% did not increase the coverage, but significantly decreased accuracy (data not shown).

Comparison with the complete collection of known complexes (279 from MIPS and the literature) showed that 257 of 491 complexes were entirely novel, and just 20 of those previously known lacked novel components (Supplementary Table S2). Of the known

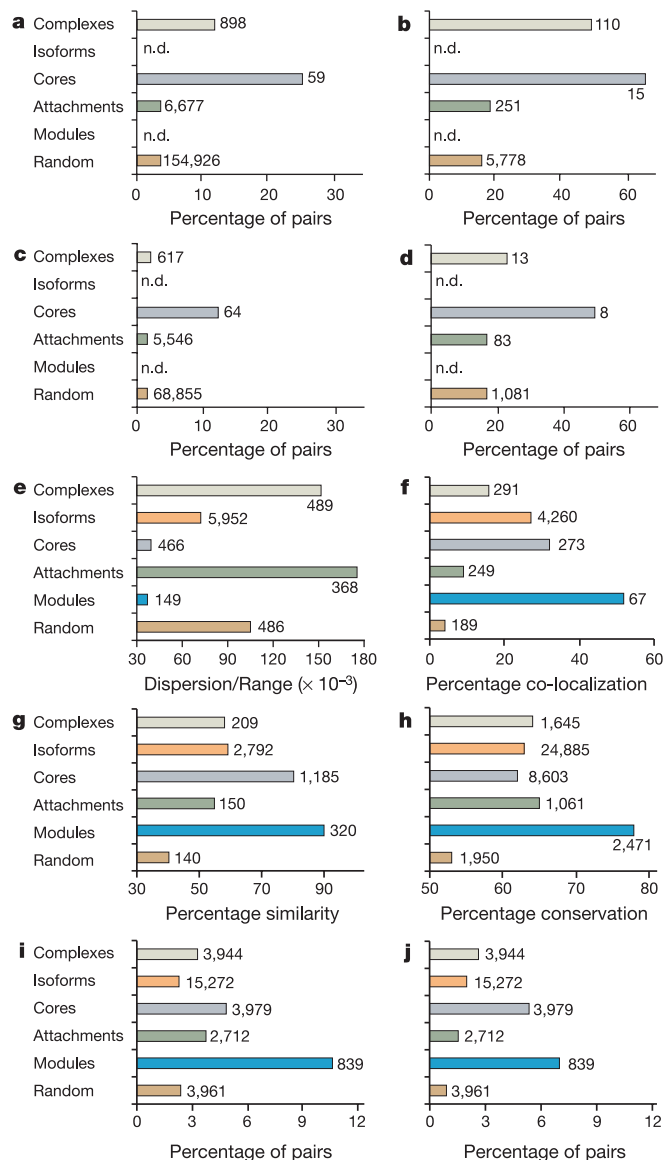


Figure 2 | Evidence supporting complex organization. Proteins in each organization level (cores, and so on) are referred to as groups. **a**, Percentage of cell cycle co-regulated genes found in the same group. **b**, Percentage of co-regulated proteins in the same group expressed at the same time during the cell cycle. **c**, **d**, are as for **a**, **b**, but for sporulation genes. **e**, Average dispersion ranges for protein abundance within each group. **f–h**, Percentage of groups having exactly the same subcellular localizations, cellular functions or phylogenetic conservation, respectively. **i**, **j**, Percentage of pairs for which a direct interaction is known from three-dimensional structures or yeast two-hybrid experiments, respectively. Values on each bar show the total number of counts; n.d., not determined. See Supplementary Information for further details.

complexes not recovered by the procedure above, 36 were partially found in single purifications (Supplementary Table S4) but produced a signal too weak to be recovered automatically.

Modular organization of the cell machinery

The above procedure partitions proteins in complexes into two types: core components that are present in most isoforms, and attachments present in only some of them (Fig. 1). This is reminiscent of an organization structure proposed previously that was based on a small-scale analysis²⁷. Complex cores ranged from 1–23 proteins in size (average 3.1 ± 2.5). Among the attachments, we noticed several instances where two or more proteins were always together and present in multiple complexes, which we call ‘modules’ (Supplementary Table S3; on average, associated with 3.3 ± 1.6 cores).

We tested whether this organization was a reflection of biological phenomena by first looking at transcriptional control of the complex components. A quality controlled set of 975 differentially expressed genes derived from microarray analyses¹⁵ showed that a large percentage of pairs of proteins within cores were coexpressed at the

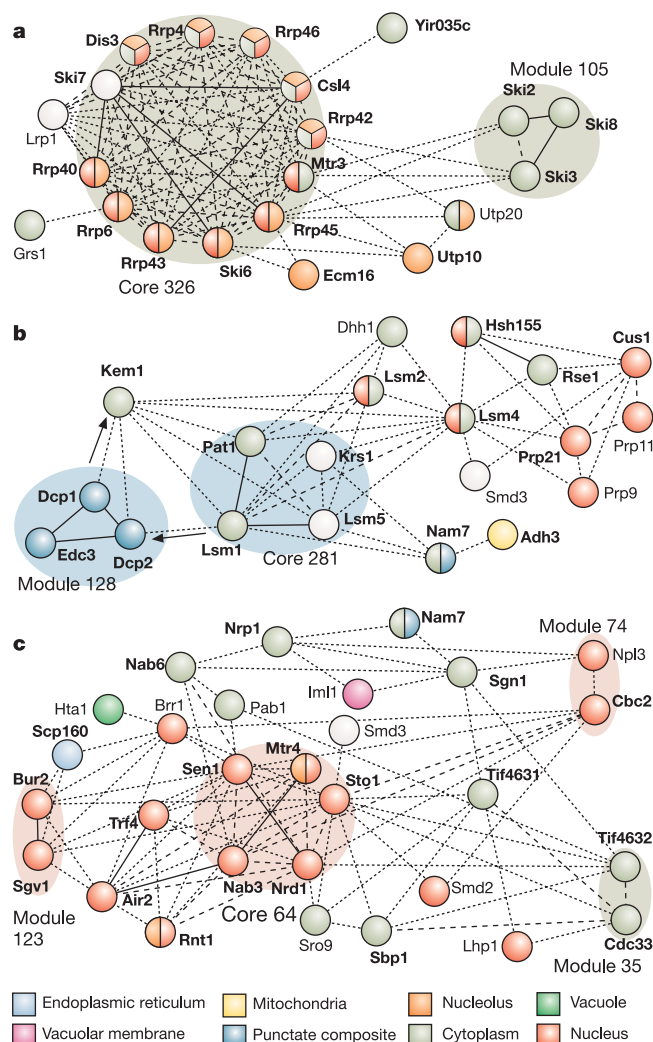


Figure 3 | Architecture and modularity of complexes. Proteins are coloured according to their localization²⁰. The line attribute corresponds to socio-affinity indices: dotted lines, 5–10; dashed lines, 10–15; plain lines, >15. Bait proteins are shown in bold and shaded circles around groups of proteins indicate cores and modules. **a**, The exosome and the Ski module. **b**, Stages in de-adenylation-dependent mRNA degradation; arrows show the order of events. **c**, Two distinct families of cap-binding proteins: the nuclear CBC (cap-binding complex) and the cytoplasmic eIF4F.

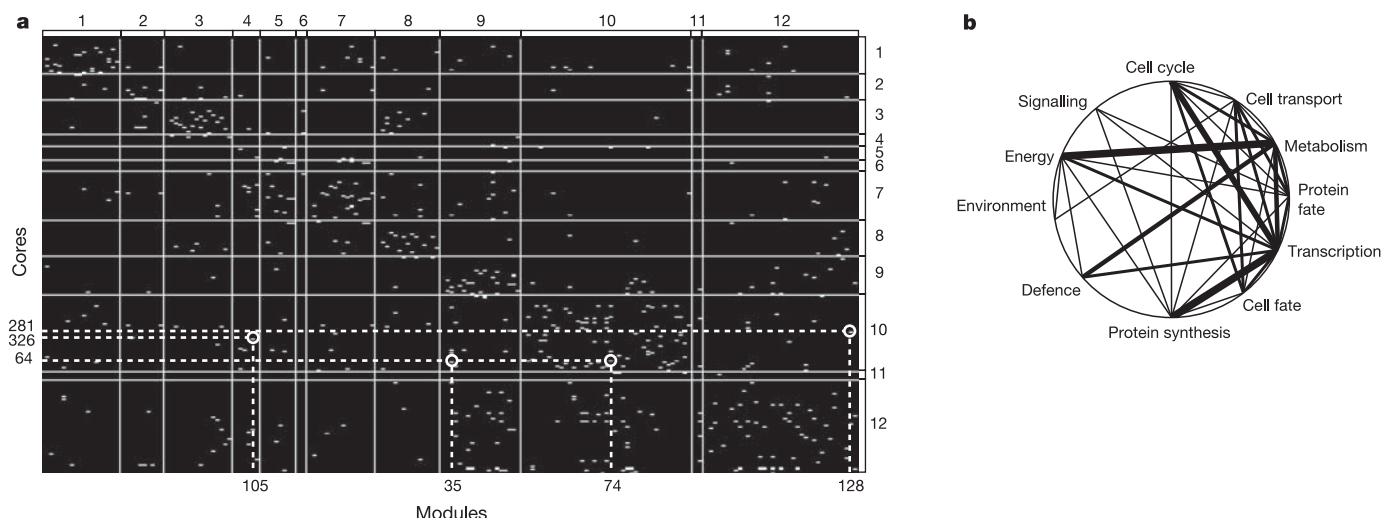


Figure 4 | Modularity of the yeast cellular machinery. **a**, Modularity matrix across cellular function. The *x* and *y* axes show modules and cores, respectively, clustered according to functional categories (1–12): cell cycle, cell fate, cell transport, defence, energy, environment, metabolism, protein fate, protein synthesis, transcription, signalling and unknown. Whenever a

module combines with a core the intersection is highlighted. Dotted lines show the modularity of the complexes in Fig. 3. **b**, Frequency of cross-talk between different cellular processes. The thickness of the lines between the functional classes are proportional to the frequency of core–module interactions between them.

same time during the cell cycle and sporulation (Fig. 2a–d), consistent with the view that core components represent functional units. Comparison with genome-wide protein abundance and localization studies^{19,20} revealed that cores and modules were also more likely to be expressed at a similar copy number (Fig. 2e) and to be co-localized in the cell (Fig. 2f). Notably, attachments showed a greater heterogeneity in expression levels than expected from random, supporting the notion that they might represent non-stoichiometric components. Cores and modules showed the greatest degree of similarity in terms of annotated function (Fig. 2g). When considering orthologous proteins in other species, cores and modules were least likely to be present partially: that is, if one component was present (or absent), the others usually were also (Fig. 2h). Finally, proteins within cores and modules were most likely to be in direct physical contact, as assessed both by three-dimensional structures (Fig. 2i) and the yeast two-hybrid system (Fig. 2j). Overall, the greatest degree of functional similarity and physical association was found between proteins within cores or modules, thus strongly supporting the model.

Examples of protein-complex architecture

The analysis was able to capture architectural details of known complexes. Attachments often specify a particular function for a complex. The exosome contains the complete Ski complex among its attachments (Fig. 3a), supporting previous reports that this association is required for cytoplasmic messenger RNA 3′-to-5′ decay²⁸. The modular architecture can also capture sequential events associated with pathways, providing a dynamic view of cellular processes. Complex 281 captured three discrete functional stages in de-adenylation-dependent RNA degradation (Fig. 3b). The core of the complex binds to de-adenylated mRNAs, a module (Edc3–Dcp1–Dcp2; known as the mRNA de-capping complex) removes the 5′ cap, and the attachment protein Kem1 (a 5′-3′ exonuclease) digests the RNA²⁹.

We identified 87 mutually exclusive modules in 48 complexes. Of these, 31 appeared to be related to differences in subcellular locations and might thus specify subtle differences in function. Among them, two mutually exclusive cap-binding modules were in different isoforms of complex 64 (Fig. 3c). The first, Tif4632–Cdc33 (or eIF4F), is cytoplasmic and essential for cap-dependent translation, while the second is nuclear and plays a direct role in pre-mRNA processing and export^{30,31}.

Other architectures hinted at novel regulatory mechanisms. Complex 437, formed around the yeast 14-3-3 protein Bmh2,

contained three metabolic enzymes involved in the heat stress response³²: Nth1, a neutral trehalase and the serine palmitoyl-transferase complex Lcb1–Lcb2. Nth1 contained three predicted 14-3-3-binding motifs and formed a core with Bmh2. The presence of Lcb1–Lcb2 as a module suggested the assembly of alternative complexes around Bmh2. A common control mechanism for Nth1 and Lcb1–Lcb2 might ensure the coordinated production of two metabolites central to the heat shock response—trehalose and sphingolipids. Similar coordinated control of metabolic enzymes through phosphorylation and subsequent binding to 14-3-3 is established in plants³³ and has recently been proposed for human cells³⁴.

A modularity matrix across functions

We derived a matrix representing a global view of the connections between cores and modules (Fig. 4a). There was a strong tendency for modules to combine with cores in the same functional category, suggesting coherence in our assignment of core and module composition. Using the ‘guilt-by-association’ principle, it is possible to suggest functions for modules. For example, the novel module 78 (Kre33 and Ygr145w) combined with several cores involved in ribosome biogenesis, suggesting a role in this process. Module 115 (Sgn1 and Ygr250c) associated with the translation initiation complex eIF4G, supporting previous genetic evidence for a role in RNA metabolism³⁵.

The degree of core–module cross-talk between functional categories (Fig. 4b) highlights many known connections, such as that between protein synthesis, transcription and the cell cycle, in addition to others less well established. For instance, the many links between metabolism and transcription are supported by recent findings of roles for metabolic enzymes in transcriptional regulation³⁶. Similarly, strong links between cell metabolism and defence argue for a re-evaluation of yeast metabolic pathways as targets for anti-fungal drug discovery.

Complexes as a scaffold for genetic data

Interaction networks have been used previously to study the effect of gene knockouts, for example showing that proteins central in networks tend to be lethal when deleted³⁷. More recently, studies have systematically monitored the effects of loss of function under a series of different conditions^{38,39} leading to phenotypic profiles, which are ideal for probing protein-complex architecture (Fig. 5). We found 20

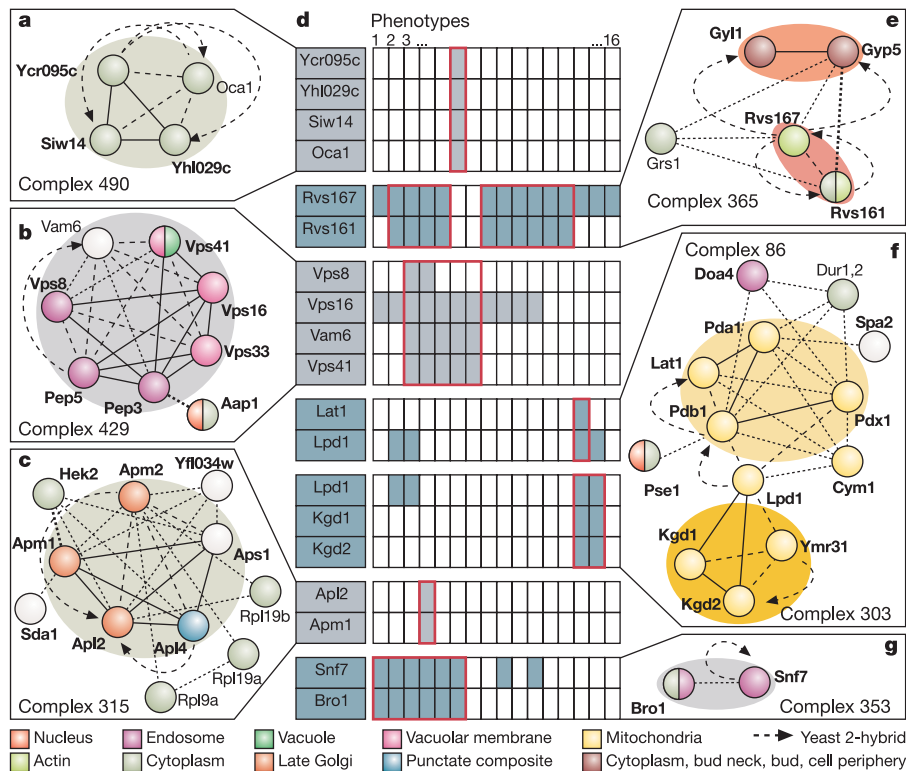


Figure 5 | Phenotypic data mapped to complexes. **a**, Novel complex 490; **b**, HOPS (homotypic fusion and vacuole protein sorting) complex⁴¹; **c**, AP1 adaptor complex; **e**, Rvs161–Rvs167 amphiphysin-like complex and the module Gyl1–Gyp5⁴²; **f**, Pyruvate and α-ketoglutarate dehydrogenase complexes⁴³; **g**, Bro1–Snf7 complex. Details are as for Fig. 3. **d**, Phenotypic effect of deletion of complex components³⁸. Shaded cells indicate a growth defect (slow growth or no growth relative to the control); those boxed in red

represent the phenotypic signature of the complex. Similarities (mean number of phenotypes shared by components/total number of phenotypes) were calculated for 20 complexes. Sensitivity phenotypes (1–16): paraquat, ethanol, CdCl₂, hygromycin-B, CaCl₂, caffeine, rapamycin, cycloheximide, hydroxyurea, galactose, high salt, raffinose, glycerol, lactate, benomyl and low phosphate.

complexes with at least two proteins present in a data set of yeast phenotypes³⁸, of which 16 showed similar phenotypic patterns (Fig. 5d; random behaviour would predict only five). In one case, profile similarity supported the authenticity of a novel complex (Fig. 5a). In others, there is evidence that shared proteins play wider roles than the individual complexes they are part of. For example, the pyruvate and α-ketoglutarate dehydrogenase complexes show similar phenotypes, but the lipoamide dehydrogenase subunit (Lpd1) shared between them has other phenotypes, suggesting that it could have additional functions (Fig. 5f). These examples highlight the promise for the molecular machinery described here to provide a molecular rationale for gene-to-phenotype relationships.

Discussion

This analysis represents only a snapshot of the proteome averaged over all phases of the cell cycle. Nevertheless, this is the first screen for complexes run to saturation and, as such, it serves as a guide for the future exploration of protein interactions under other physiological states. For example, we do not expect protein-complex cores to vary extensively under different conditions, whereas we expect significant changes to occur in attachment proteins. Extrapolation based on the fraction of known complexes recovered suggests that there may be an additional 300 core machines, leading to a total of 800 in yeast. In a rough approximation, based on the ratio of gene numbers between species, we estimate some 3,000 core human complexes.

The number of protein-complex cores is small compared to the many cellular processes mediated by them, and shuffling functional modules provides an efficient means to multiply functionality and simplify temporal and spatial regulation. The modularity is highly reminiscent of that seen elsewhere in nature, for example the combinatorial use of amino acids to build polypeptides, or domains

to create proteins with complex biochemical properties. Modularity might very well represent a general attribute of living matter, with *de novo* invention being rare and reuse the norm.

Genome sequencing and functional genomics have provided a parts-list and partial knowledge of how these parts are arranged in space and time. The next challenge is to integrate these data into rational models of entire systems. Our analysis makes some first steps in this direction, providing a collection of individual integrative subsystems—the machines—but also a view on how they might coordinate cellular functions through sharing functional modules. As such, it may be a very useful platform for systems biology and indeed new applications in nano- and synthetic-biology that seek to re-engineer the cellular machinery towards new processes.

METHODS

Experimental procedures. We created a library of strains with TAP-tag cassettes at the 3' end of each ORF by homologous recombination. We prepared protein extracts from exponentially growing haploid yeast strains grown in 21 of complete medium. Tandem-affinity purification (TAP)–mass spectrometry (MS) characterization of complexes was performed as previously described⁸. For membrane proteins, we used a special protocol provided as Supplementary Information.

Socio-affinity and iterative clustering to generate protein-complex sets. We defined a socio-affinity index ($A(i,j)$) that quantifies the tendency for proteins to identify each other when tagged (the spoke model, S) and to co-purify when other proteins are tagged (the matrix model, M)⁴⁰:

$$A(i,j) = S_{i,j|i=\text{bait}} + S_{i,j|j=\text{bait}} + M_{i,j}; \quad S_{i,j|i=\text{bait}} = \log \left(\frac{n_{i,j|i=\text{bait}}}{f_i^{\text{bait}} n_{\text{bait}} f_j^{\text{prey}} n_{i=\text{bait}}^{\text{prey}}} \right)$$

$$M_{i,j} = \log \left(\frac{n_{i,j}^{\text{prey}}}{f_i^{\text{prey}} f_j^{\text{prey}} \sum_{\text{all baits}} n_{\text{prey}} (n_{\text{prey}} - 1)/2} \right)$$

For the spoke model terms (S), $n_{i,j|i=bait}$ is the number of times that protein i retrieves j when i is tagged; f_i^{bait} is the fraction of purifications where protein i was bait; f_j^{prey} is the fraction of all retrieved preys that were protein j ; n_{bait} is the total number of purifications (that is, baits); and $n_{i=bait}^{prey}$ is the number of preys retrieved with protein i as bait. For the matrix model term (M), $n_{i,j}^{prey}$ is the number of times that proteins i and j are seen in purifications with baits other than i or j ; f_i^{prey} and f_j^{prey} are as above; and n_{prey} is the number of preys observed with a particular bait (excluding itself).

We used socio-affinity indices to populate the upper-diagonal of a pair-wise matrix (that is, one value for each pair of proteins in the data set). We assigned a value of zero to all pairs of proteins that had never been seen together. We generated a first set of clusters using the OC program (G. Barton, University of Dundee) and then subtracted a penalty from each pair-wise value associated with the set. We then repeated the cluster generation a number of times, each time adding any new clusters to a growing list. To generate different sets of complexes using this procedure, we varied the number of iterations (2–10), the socio-affinity threshold to define clusters (1–10), the penalty value (0.5, 1 or 2), and the type of clustering (UPGMA, single or complete linkage).

Received 17 October; accepted 15 December 2005.

Published online 22 January 2006.

- Hood, L., Heath, J. R., Phelps, M. E. & Lin, B. Systems biology and new technologies enable predictive and preventative medicine. *Science* **306**, 640–643 (2004).
- Alberts, B. The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell* **92**, 291–294 (1998).
- Goh, C. S., Milburn, D. & Gerstein, M. Conformational changes associated with protein–protein interactions. *Curr. Opin. Struct. Biol.* **14**, 104–109 (2004).
- Kemmeren, P. *et al.* Protein interaction verification and functional annotation by integrated analysis of genome-scale data. *Mol. Cell* **9**, 1133–1143 (2002).
- Edwards, A. M. *et al.* Bridging structural biology and genomics: assessing protein interaction data with known complexes. *Trends Genet.* **18**, 529–536 (2002).
- Puig, O. *et al.* The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods* **24**, 218–229 (2001).
- Rigaut, G. *et al.* A generic protein purification method for protein complex characterization and proteome exploration. *Nature Biotechnol.* **17**, 1030–1032 (1999).
- Gavin, A. C. *et al.* Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141–147 (2002).
- Ho, Y. *et al.* Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* **415**, 180–183 (2002).
- Bouwmeester, T. *et al.* A physical and functional map of the human TNF- α /NF- κ B signal transduction pathway. *Nature Cell Biol.* **6**, 97–105 (2004).
- Butland, G. *et al.* Interaction network containing conserved and essential protein complexes in *Escherichia coli*. *Nature* **433**, 531–537 (2005).
- Uetz, P. *et al.* A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. *Nature* **403**, 623–627 (2000).
- Rual, J. F. *et al.* Towards a proteome-scale map of the human protein–protein interaction network. *Nature* **437**, 1173–1178 (2005).
- Aloy, P. *et al.* Structure-based assembly of protein complexes in yeast. *Science* **303**, 2026–2029 (2004).
- de Lichtenberg, U., Jensen, L. J., Brunak, S. & Bork, P. Dynamic complex formation during the yeast cell cycle. *Science* **307**, 724–727 (2005).
- Kelley, R. & Ideker, T. Systematic interpretation of genetic interactions using protein networks. *Nature Biotechnol.* **23**, 561–566 (2005).
- Mewes, H. W. *et al.* MIPS: A database for genomes and protein sequences. *Nucleic Acids Res.* **30**, 31–34 (2002).
- Kumar, A. *et al.* An integrated approach for finding overlooked genes in yeast. *Nature Biotechnol.* **20**, 58–63 (2002).
- Ghaemmaghami, S. *et al.* Global analysis of protein expression in yeast. *Nature* **425**, 737–741 (2003).
- Huh, W. K. *et al.* Global analysis of protein localization in budding yeast. *Nature* **425**, 686–691 (2003).
- Washburn, M. P., Wolters, D. & Yates, J. R. III Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnol.* **19**, 242–247 (2001).
- Mewes, H. W. *et al.* MIPS: Analysis and annotation of proteins from whole genomes. *Nucleic Acids Res.* **32**, D41–D44 (2004).
- Liou, A. K. & Willison, K. R. Elucidation of the subunit orientation in CCT (chaperonin containing TCP1) from the subunit composition of CCT micro-complexes. *EMBO J.* **16**, 4311–4316 (1997).
- Kraynack, B. A. *et al.* Dsl1p, Tip20p, and the novel Dsl3(Sec39) protein are required for the stability of the Q/t-SNARE complex at the endoplasmic reticulum in yeast. *Mol. Biol. Cell* **16**, 3963–3977 (2005).
- Kao, L. R., Peterson, J., Ji, R., Bender, L. & Bender, A. Interactions between the ankyrin repeat-containing protein Akrlp and the pheromone response pathway in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **16**, 168–178 (1996).
- Bader, G. D. & Hogue, C. W. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* **4**, 2 (2003).
- Dezso, Z., Oltvai, Z. N. & Barabasi, A. L. Bioinformatics analysis of experimentally determined protein complexes in the yeast *Saccharomyces cerevisiae*. *Genome Res.* **13**, 2450–2454 (2003).
- Orban, T. I. & Izaurralde, E. Decay of mRNAs targeted by RISC requires XRN1, the Ski complex, and the exosome. *RNA* **11**, 459–469 (2005).
- Sheth, U. & Parker, R. Decapping and decay of messenger RNA occur in cytoplasmic processing bodies. *Science* **300**, 805–808 (2003).
- Fortes, P. *et al.* The yeast nuclear cap binding complex can interact with translation factor eIF4G and mediate translation initiation. *Mol. Cell* **6**, 191–196 (2000).
- McKendrick, L., Thompson, E., Ferreira, J., Morley, S. J. & Lewis, J. D. Interaction of eukaryotic translation initiation factor 4G with the nuclear cap-binding complex provides a link between nuclear and cytoplasmic functions of the m⁷ guanosine cap. *Mol. Cell. Biol.* **21**, 3632–3641 (2001).
- Jenkins, G. M. & Hannun, Y. A. Role for *de novo* sphingoid base biosynthesis in the heat-induced transient cell cycle arrest of *Saccharomyces cerevisiae*. *J. Biol. Chem.* **276**, 8574–8581 (2001).
- Sehnke, P. C. & Ferl, R. J. Plant 14-3-3s: Omnipotent metabolic phosphopartners? *Sci. STKE* **2000**, PE1 (2000) (doi:10.1126/stke.2000.56.pe1).
- Pozuelo Rubio, M. *et al.* 14-3-3-affinity purification of over 200 human phosphoproteins reveals new links to regulation of cellular metabolism, proliferation and trafficking. *Biochem. J.* **379**, 395–408 (2004).
- Winstall, E., Sadowski, M., Kuhn, U., Wahle, E. & Sachs, A. B. The *Saccharomyces cerevisiae* RNA-binding protein Rbp29 functions in cytoplasmic mRNA metabolism. *J. Biol. Chem.* **275**, 21817–21826 (2000).
- Shi, Y. & Shi, Y. Metabolic enzymes and coenzymes in transcription—a direct link between metabolism and transcription? *Trends Genet.* **20**, 445–452 (2004).
- Jeong, H., Mason, S. P., Barabasi, A. L. & Oltvai, Z. N. Lethality and centrality in protein networks. *Nature* **411**, 41–42 (2001).
- Dudley, A. M., Janse, D. M., Tanay, A., Shamir, R. & Church, G. M. A global view of pleiotropy and phenotypically derived gene function in yeast. *Mol. Syst. Biol.* published online 29 March 2005 (doi:10.1038/msb4100004).
- Said, M. R., Begley, T. J., Oppenheim, A. V., Lauffenburger, D. A. & Samson, L. D. Global network analysis of phenotypic effects: protein networks and toxicity modulation in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* **101**, 18006–18011 (2004).
- Bader, G. D. & Hogue, C. W. Analyzing yeast protein–protein interaction data obtained from different sources. *Nature Biotechnol.* **20**, 991–997 (2002).
- Subramanian, S., Woolford, C. A. & Jones, E. W. The Sec1/Munc18 protein, Vps33p, functions at the endosome and the vacuole of *Saccharomyces cerevisiae*. *Mol. Biol. Cell* **15**, 2593–2605 (2004).
- Friesen, H., Colwill, K., Robertson, K., Schub, O. & Andrews, B. Interaction of the *Saccharomyces cerevisiae* cortical actin patch protein Rvs167p with proteins involved in ER to golgi vesicle trafficking. *Genetics* **170**, 555–568 (2005).
- Ross, J., Reid, G. A. & Dawes, I. W. The nucleotide sequence of the *LPD1* gene encoding lipamide dehydrogenase in *Saccharomyces cerevisiae*: comparison between eukaryotic and prokaryotic sequences for related enzymes and identification of potential upstream control sites. *J. Gen. Microbiol.* **134**, 1131–1139 (1988).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Cohen, S. Artavanis-Tsakonas, B. Seraphin and L. Serrano for support and suggestions throughout the work, and F. Weisbrodt for assistance with the graphics.

Author Information Purification and complex data have been deposited at the IntAct database (<http://www.ebi.ac.uk/intact/>) with accession numbers EBI-768904 (purifications) and EBI-765905 (author inferred complexes). The data, including the MS protein identifications, are accessible at <http://yeast-complexes.embl.de>, and the yeast strains are available from Euroscarf (http://web.uni-frankfurt.de/fb15/mikro/euroscarf/col_index.html). Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to G.S.-F. (gsuperti@cemm.oeaw.ac.at), R.B.R. (russell@embl-heidelberg.de).

Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*

Nevan J. Krogan^{1,2,*†}, Gerard Cagney^{1,3,*}, Haiyuan Yu⁴, Gouqing Zhong¹, Xinghua Guo¹, Alexandr Ignatchenko¹, Joyce Li¹, Shuye Pu⁵, Nira Datta¹, Aaron P. Tikuisis¹, Thanuja Punna¹, José M. Peregrín-Alvarez⁵, Michael Shales¹, Xin Zhang¹, Michael Davey¹, Mark D. Robinson¹, Alberto Paccanaro⁴, James E. Bray¹, Anthony Sheung¹, Bryan Beattie⁶, Dawn P. Richards⁶, Veronica Canadien⁶, Atanas Lalev¹, Frank Mena⁶, Peter Wong¹, Andrei Starostine¹, Myra M. Canete¹, James Vlasblom⁵, Samuel Wu⁵, Chris Orsi⁵, Sean R. Collins⁷, Shamanta Chandran¹, Robin Haw¹, Jennifer J. Rilstone¹, Kiran Gandhi¹, Natalie J. Thompson¹, Gabe Musso¹, Peter St Onge¹, Shaun Ghanny¹, Mandy H. Y. Lam^{1,2}, Gareth Butland¹, Amin M. Altaf-Ul⁸, Shigehiko Kanaya⁸, Ali Shilatifard⁹, Erin O'Shea¹⁰, Jonathan S. Weissman⁷, C. James Ingles^{1,2}, Timothy R. Hughes^{1,2}, John Parkinson⁵, Mark Gerstein⁴, Shoshana J. Wodak⁵, Andrew Emili^{1,2} & Jack F. Greenblatt^{1,2}

Identification of protein–protein interactions often provides insight into protein function, and many cellular processes are performed by stable protein complexes. We used tandem affinity purification to process 4,562 different tagged proteins of the yeast *Saccharomyces cerevisiae*. Each preparation was analysed by both matrix-assisted laser desorption/ionization–time of flight mass spectrometry and liquid chromatography tandem mass spectrometry to increase coverage and accuracy. Machine learning was used to integrate the mass spectrometry scores and assign probabilities to the protein–protein interactions. Among 4,087 different proteins identified with high confidence by mass spectrometry from 2,357 successful purifications, our core data set (median precision of 0.69) comprises 7,123 protein–protein interactions involving 2,708 proteins. A Markov clustering algorithm organized these interactions into 547 protein complexes averaging 4.9 subunits per complex, about half of them absent from the MIPS database, as well as 429 additional interactions between pairs of complexes. The data (all of which are available online) will help future studies on individual proteins as well as functional genomics and systems biology.

Elucidation of the budding yeast genome sequence¹ initiated a decade of landmark studies addressing key aspects of yeast cell biology on a system-wide level. These included microarray-based analysis of gene expression², screens for various biochemical activities^{3,4}, identification of protein subcellular locations^{5,6}, and identifying effects of single and pairwise gene disruptions^{7–10}. Other efforts were made to catalogue physical interactions among yeast proteins, primarily using the yeast two-hybrid method^{11,12} and direct purification via affinity tags^{13,14}; many of these interactions are conserved in other organisms¹⁵. Data from the yeast protein–protein interaction studies have been non-overlapping to a surprising degree, a fact explained partly by experimental inaccuracy and partly by indications that no single screen has been comprehensive¹⁶.

Proteome-wide purification of protein complexes

Of the various high throughput experimental methods used thus far to identify protein–protein interactions^{11–14}, tandem affinity purification (TAP) of affinity-tagged proteins expressed from their

natural chromosomal locations followed by mass spectrometry^{13,17} has provided the best coverage and accuracy¹⁶. To map more completely the yeast protein interaction network (interactome), *S. cerevisiae* strains were generated with in-frame insertions of TAP tags individually introduced by homologous recombination at the 3' end of each predicted open reading frame (ORF) (<http://www.yeastgenome.org/>)^{18,19}. Proteins were purified from 4L yeast cultures under native conditions, and the identities of the co-purifying proteins (preys) determined in two complementary ways¹⁷. Each purified protein preparation was electrophoresed on an SDS polyacrylamide gel, stained with silver, and visible bands removed and identified by trypsin digestion and peptide mass fingerprinting using matrix-assisted laser desorption/ionization–time of flight (MALDI–TOF) mass spectrometry. In parallel, another aliquot of each purified protein preparation was digested in solution and the peptides were separated and sequenced by data-dependent liquid chromatography tandem mass spectrometry (LC–MS/MS)^{17,20–22}. Because either mass spectrometry method often fails to

¹Banting and Best Department of Medical Research, Terrence Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, 160 College St, Toronto, Ontario M5S 3E1, Canada. ²Department of Medical Genetics and Microbiology, University of Toronto, 1 Kings College Circle, Toronto, Ontario M5S 1A8, Canada. ³Conway Institute, University College Dublin, Belfield, Dublin 4, Ireland. ⁴Department of Molecular Biophysics and Biochemistry, 266 Whitney Avenue, Yale University, PO Box 208114, New Haven, Connecticut 06520, USA. ⁵Hospital for Sick Children, 555 University Avenue, Toronto, Ontario M4K 1X8, Canada. ⁶Affinium Pharmaceuticals, 100 University Avenue, Toronto, Ontario M5J 1V6, Canada. ⁷Howard Hughes Medical Institute, Department of Cellular and Molecular Pharmacology, UCSF, Genentech Hall S472C, 600 16th St, San Francisco, California 94143, USA. ⁸Comparative Genomics Laboratory, Nara Institute of Science and Technology 8916-5, Takayama, Ikoma, Nara 630-0101, Japan. ⁹Department of Biochemistry, Saint Louis University School of Medicine, 1402 South Grand Boulevard, St Louis, Missouri 63104, USA. ¹⁰Howard Hughes Medical Institute, Department of Molecular and Cellular Biology, Harvard University, 7 Divinity Avenue, Cambridge, Massachusetts 02138, USA. [†]Present address: Department of Cellular and Molecular Pharmacology, UCSF, San Francisco, California 94143, USA.

*These authors contributed equally to this work.

identify a protein, we used two independent mass spectrometry methods to increase interactome coverage and confidence. Among the attempted purifications of 4,562 different proteins (Supplementary Table S1), including all predicted non-membrane proteins, 2,357 purifications were successful (Supplementary Table S2) in that at least one protein was identified (in 1,613 cases by MALDI-TOF mass spectrometry and in 2,001 cases by LC-MS/MS; Fig. 1a) that was not present in a control preparation from an untagged strain.

In total, 4,087 different yeast proteins were identified as preys with high confidence ($\geq 99\%$; see Methods) by MALDI-TOF mass spectrometry and/or LC-MS/MS, corresponding to 72% of the predicted yeast proteome (Supplementary Table S3). Smaller proteins with a relative molecular mass (M_r) of 35,000 were less likely to be identified (Fig. 1b), perhaps because they generate fewer peptides suited for identification by mass spectrometry. We were more successful in identifying smaller proteins by LC-MS/MS than by MALDI-TOF mass spectrometry, probably because smaller proteins stain less well with silver or ran off the SDS gels. Our success in protein identification was unrelated to protein essentiality (data not shown) and ranged from 80% for low abundance proteins to over 90% for high abundance proteins (Fig. 1c). Notably, we identified 47% of the proteins not detected by genome-wide western blotting¹⁸, indicating that affinity purification followed by mass spectrometry can be more sensitive. Many hypothetical proteins not detected by western blotting¹⁸ or our mass spectrometry analyses may not be expressed in our standard cell growth conditions. Although our success rates for identifying proteins were 94% and 89% for nuclear and cytosolic proteins, respectively, and at least 70% in most cellular compartments (Fig. 1d), they were lower (61% and 59%, respectively) for the endoplasmic reticulum and vacuole. However, even though we had not tagged or purified most proteins with transmembrane

domains, we identified over 70% of the membrane-associated proteins, perhaps because our extraction and purification buffers contained 0.1% Triton X-100. Our identification success rate was lowest (49%) with proteins for which localization was not established^{15,6}, many of which may not be expressed. We had high success in identifying proteins involved in all biological processes, as defined by gene ontology (GO) nomenclature, or possessing any broadly defined GO molecular function (Fig. 1e, f). We were less successful (each about 65% success) with transporters and proteins of unknown function; many of the latter may not be expressed.

A high-quality data set of protein-protein interactions

Deciding whether any two proteins interact based on our data must encompass results from two purifications (plus repeat purifications, if performed) and integrate reliability scores from all protein identifications by mass spectrometry. Removed from consideration as likely nonspecific contaminants were 44 preys detected in $\geq 3\%$ of the purifications and nearly all cytoplasmic ribosomal subunits (Supplementary Table S4). Although the cytosolic ribosomes and pre-ribosomes, as well as some associated translation factors, are not represented in the interaction network and protein complexes we subsequently identified, we previously described the interactome for proteins involved in RNA metabolism and ribosome biogenesis²².

We initially generated an 'intersection data set' of 2,357 protein-protein interactions based only on proteins identified in at least one purification by both MALDI-TOF mass spectrometry and LC-MS/MS with relatively low thresholds (70%) (Supplementary Table S5). This intersection data set containing 1,210 proteins was of reasonable quality but limited in scope (Fig. 2b). Our second approach added to the intersection data set proteins identified either reciprocally or repeatedly by only a single mass spectrometry method

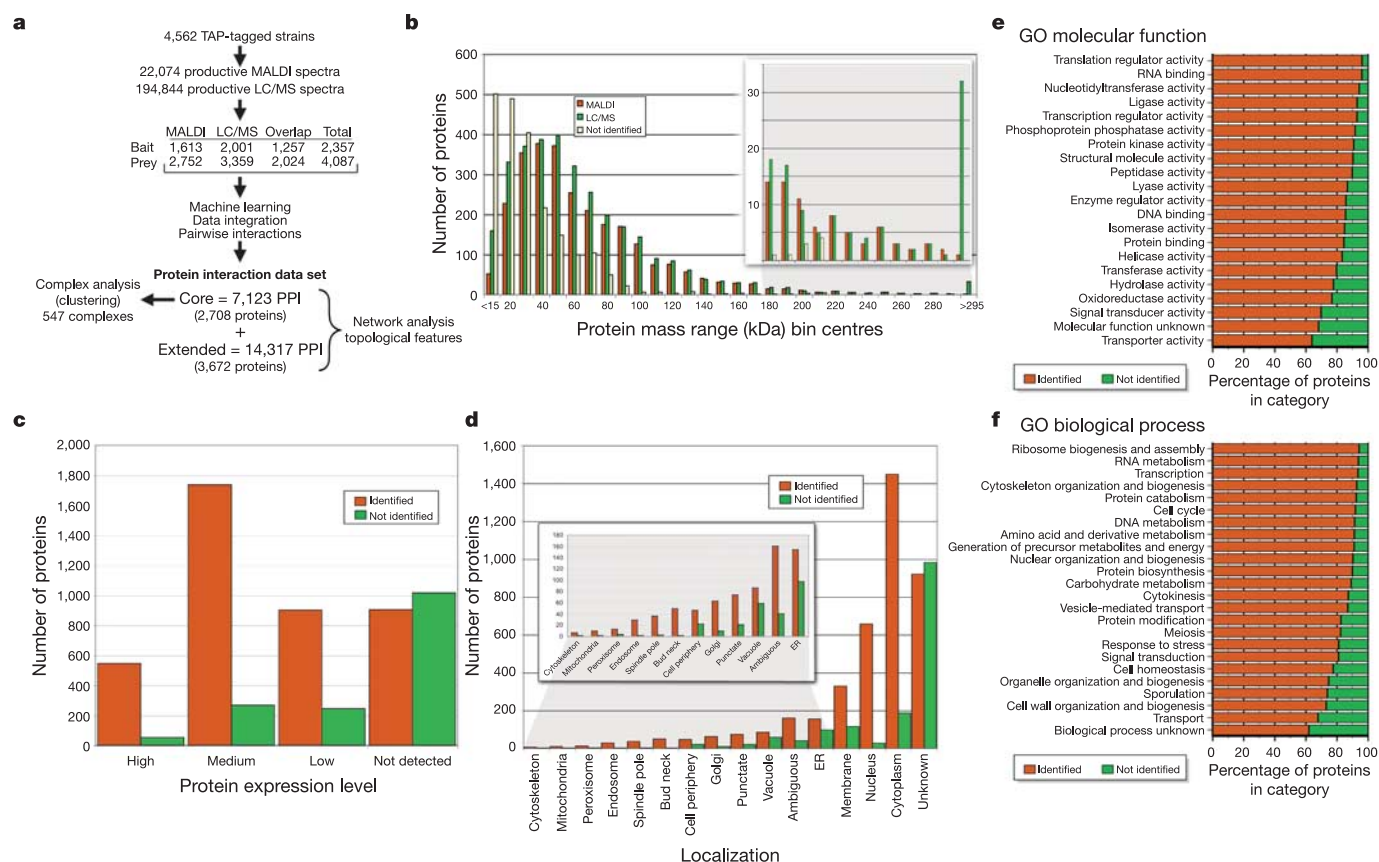


Figure 1 | The yeast interactome encompasses a large proportion of the predicted proteome. a, Summary of our screen for protein interactions. PPI, protein-protein interactions. **b–f**, The proportions of proteins

identified in the screen as baits or preys are shown in relation to protein mass (**b**), expression level (**c**), intracellular localization (**d**) and annotated GO molecular function (**e**) and GO biological process (**f**).

to generate the 'merged data set'. The merged data set containing 2,186 proteins and 5,496 protein–protein interactions (Supplementary Table S6) had better coverage than the intersection network (Fig. 2b).

To deal objectively with noise in the raw data and improve precision and recall, we used machine learning algorithms with two rounds of learning. All four classifiers were validated by the hold-out method (66% for training and 33% for testing) and ten-times tenfold cross-validation, which gave similar results. Because our objective was to identify protein complexes, we used the hand-curated protein complexes in the MIPS reference database²³ as our training set. Our goal was to assign a probability that each pairwise interaction is true based on experimental reproducibility and mass spectrometry scores from the relevant purifications (see Methods). In the first round of learning, we tested bayesian inference networks and 28 different kinds of decision trees²⁴, settling on bayesian networks and C4.5-based and boosted stump decision trees as providing the most reliable predictions (Fig. 2a). We then improved performance by using the output of the three methods as input for a second round of learning with a stacking algorithm in which logistic regression was the learner²⁵. We used a probability cut-off of 0.273 (average 0.68; median 0.69) to define a 'core' data set of 7,123 protein–protein interactions involving 2,708 proteins (Supplementary Table S7) and a cut-off of 0.101 (average 0.42; median 0.27) for an 'extended' data set of 14,317 protein–protein interactions involving 3,672 proteins (Supplementary Table S8). The interaction probabilities in Supplementary Tables S7 and S8 are likely to be underestimated because the MIPS complexes used as a 'gold standard' are themselves imperfect²⁶. We subsequently used the core protein–protein interaction data set to define protein complexes (see below), but the extended data set probably contains at least 1,000 correct interactions (as well as many more false interactions) not present in the core data set.

The complete set of protein–protein interactions and their associated probabilities (Supplementary Table S9) were used to generate a ROC curve with a performance (area under the curve) of 0.95 (Fig. 2b). Predictive sensitivity (true positive rate) or specificity (false positive rate), or both, are superior for our learned data set than for the intersection and merged data sets, each previous high-throughput study of yeast protein–protein interactions^{11–14}, or a bayesian combination of the data from all these studies²⁷ (Fig. 2b).

Identification of complexes within the interaction network

In the protein interaction network generated by our core data set of 7,123 protein–protein interactions, the average degree (number of

interactions per protein) is 5.26 and the distribution of the number of interactions per protein follows an inverse power law (Fig. 2c), indicating scale-free network topology²⁸. These protein–protein interactions could be represented as a weighted graph (not shown) in which individual proteins are nodes and the weight of the arc connecting two nodes is the probability that interaction is correct. Because the 2,357 successful purifications underlying such a graph would represent >50% of the detectably expressed proteome¹⁸, we have typically purified multiple subunits of a given complex. To identify highly connected modules within the global protein–protein interaction network, we used the Markov cluster algorithm, which simulates random walks within graphs²⁹. We chose values for the expansion and inflation operators of the Markov cluster procedure that optimized overlap with the hand-curated MIPS complexes²³. Although the Markov cluster algorithm displays good convergence and robustness, it does not necessarily separate two or more complexes that have shared subunits (for example, RNA polymerases I and III, or chromatin modifying complexes Rpd3C(S) and Rpd3C(L))^{30,31}.

The Markov cluster procedure identified 547 distinct (non-overlapping) heteromeric protein complexes (Supplementary Table S10), about half of which are not present in MIPS or two previous high-throughput studies of yeast complexes using affinity purification and mass spectrometry (Fig. 3a). New subunits or interacting proteins were identified for most complexes that had been identified previously (Fig. 3a). Overlap of our Markov-cluster-computed complexes with the MIPS complexes was evaluated (see Supplementary Information) by calculating the total precision (measure of the extent to which proteins belonging to one reference MIPS complex are grouped within one of our complexes, and vice versa) and homogeneity (measure of the extent to which proteins from the same MIPS complex are distributed across our complexes, and vice versa) (Fig. 3b). Both precision and homogeneity were higher for the complexes generated in this study—even for the extended set of protein–protein interactions—than for complexes generated by both previous high-throughput studies of yeast complexes, perhaps because the increased number of successful purifications in this study increased the density of connections within most modules. The average number of different proteins per complex is 4.9, but the distribution (Fig. 3c), which follows an inverse power law, is characterized by a large number of small complexes, most often containing only two to four different polypeptides, and a much smaller number of very large complexes.

Proteins in the same complex should have similar function and co-localize to the same subcellular compartment. To evaluate this, we

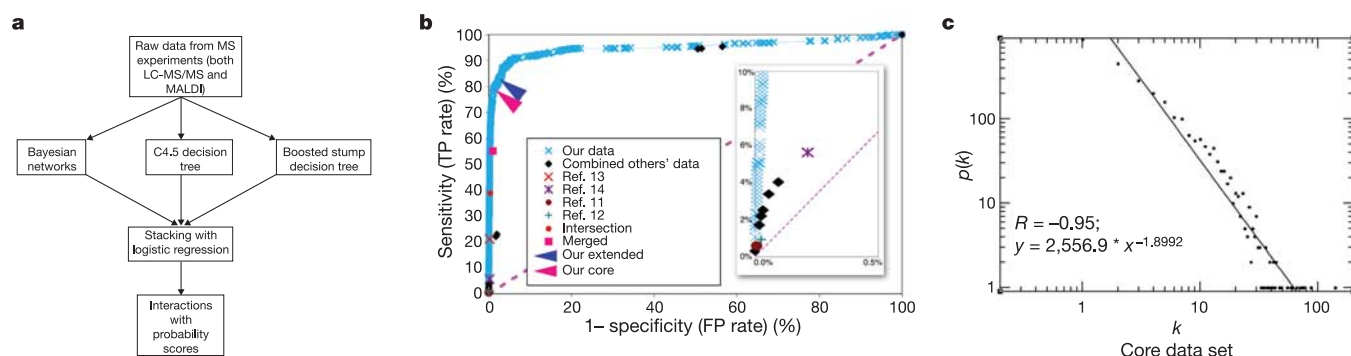
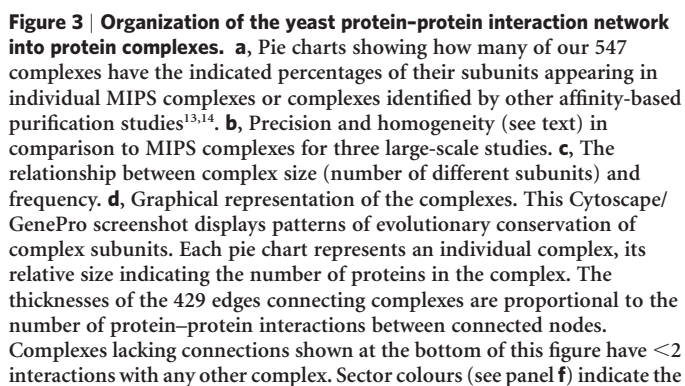


Figure 2 | Machine learning generates a core data set of protein–protein interactions. **a**, Reliability of observed protein–protein interactions was estimated using probabilistic mass spectra database search scores and measures of experimental reproducibility (see Methods), followed by machine learning. **b**, Precision–sensitivity ROC plot for our protein–protein interaction data set generated by machine learning. Precision/sensitivity values are also shown for the 'intersection' and 'merged' data sets (see text)

and for other large-scale affinity tagging^{13,14} and two-hybrid^{11,12} data sets, and a bayesian networks combination of those data sets²⁷, all based on comparison to MIPS complexes. FP, false positive; TP, true positive. **c**, Plot of the number of nodes against the number of edges per node demonstrates that the core data set protein–protein interaction network has scale-free properties.



proportion of subunits sharing significant sequence similarity to various taxonomic groups (see Methods). Insets provide views of two selected complexes—the kinetochore machinery and a previously uncharacterized, highly conserved fructose-1,6-bisphosphatase-degrading complex (see text for details)—detailing specific interactions between proteins identified within the complex (purple borders) and with other proteins that interact with at least one member of the complex (blue borders). Colours indicate taxonomic similarity. **e**, Relationship between protein frequency in the core data set and degree of connectivity or betweenness as a function of conservation. Colours of the bars indicate the evolutionary grouping. **f**, Colour key indicating the taxonomic groupings (and their phylogenetic relationships). Numbers indicate the total number of ORFs sharing significant sequence similarity with a gene in at least one organism associated with that group and, importantly, not possessing similarity to any gene from more distantly related organisms.

calculated the weighted average of the fraction of proteins in each complex that maps to the same localization categories⁵ (see Supplementary Information). Co-localization was better for the complexes in our study than for previous high-throughput studies but, not unexpectedly, less than that for the curated MIPS complexes (Supplementary Fig. S1). We also evaluated the extent of semantic similarity³² for the GO terms in the 'biological process' category for pairs of interacting proteins within our complexes (Supplementary Fig. S2), and found that semantic similarity was lower for our core data set than for the MIPS complexes or the previous study using TAP tags¹³, but higher than for a study using protein overproduction¹⁴. This might be expected if the previous TAP tag study significantly influenced the semantic classifications in GO.

To analyse and visualize our entire collection of complexes, the highly connected modules identified by Markov clustering for the global core protein–protein interaction network were displayed (<http://genepro.ccb.sickkids.ca>) using our GenePro plug-in for the Cytoscape software environment³³ (Fig. 3d). Each complex is represented as a pie-chart node, and the complexes are connected by a limited number (429) of high-confidence interactions. Assignment of connecting proteins to a particular module can therefore be arbitrary, and the limited number of connecting proteins could just as well be part of two or more distinct complexes.

The size and colour of each section of a pie-chart node can be made to represent the fraction of the proteins in each complex that maps into a given complex from the hand-curated MIPS complexes (Supplementary Fig. S3). Similar displays can be generated when highlighting instead the subcellular localizations or GO biological process functional annotations of proteins in each complex. Furthermore, the protein–protein interaction details of individual complexes can readily be visualized (see Supplementary Information).

Evolutionary conservation of protein complexes

ORFs encoding each protein were placed into nine distinct evolutionary groups (Fig. 3f) based on their taxonomic profiles (see Methods), and the complexes displayed so as to show the evolutionary conservation of their components (Fig. 3d). Insets highlight the kinetochore complex required for chromosome segregation and a novel, highly conserved complex involved in degradation of fructose-1,6-bisphosphatase. Strong co-evolution was evident for components of some large and essential complexes (for example, 19S and 20S proteasomes involved in protein degradation, the exosome involved in RNA metabolism, and the ARP2/3 complex required for the motility and integrity of cortical actin patches). Conversely, the kinetochore complex, the mediator complex required for regulated transcription, and the RSC complex that remodels chromatin have a

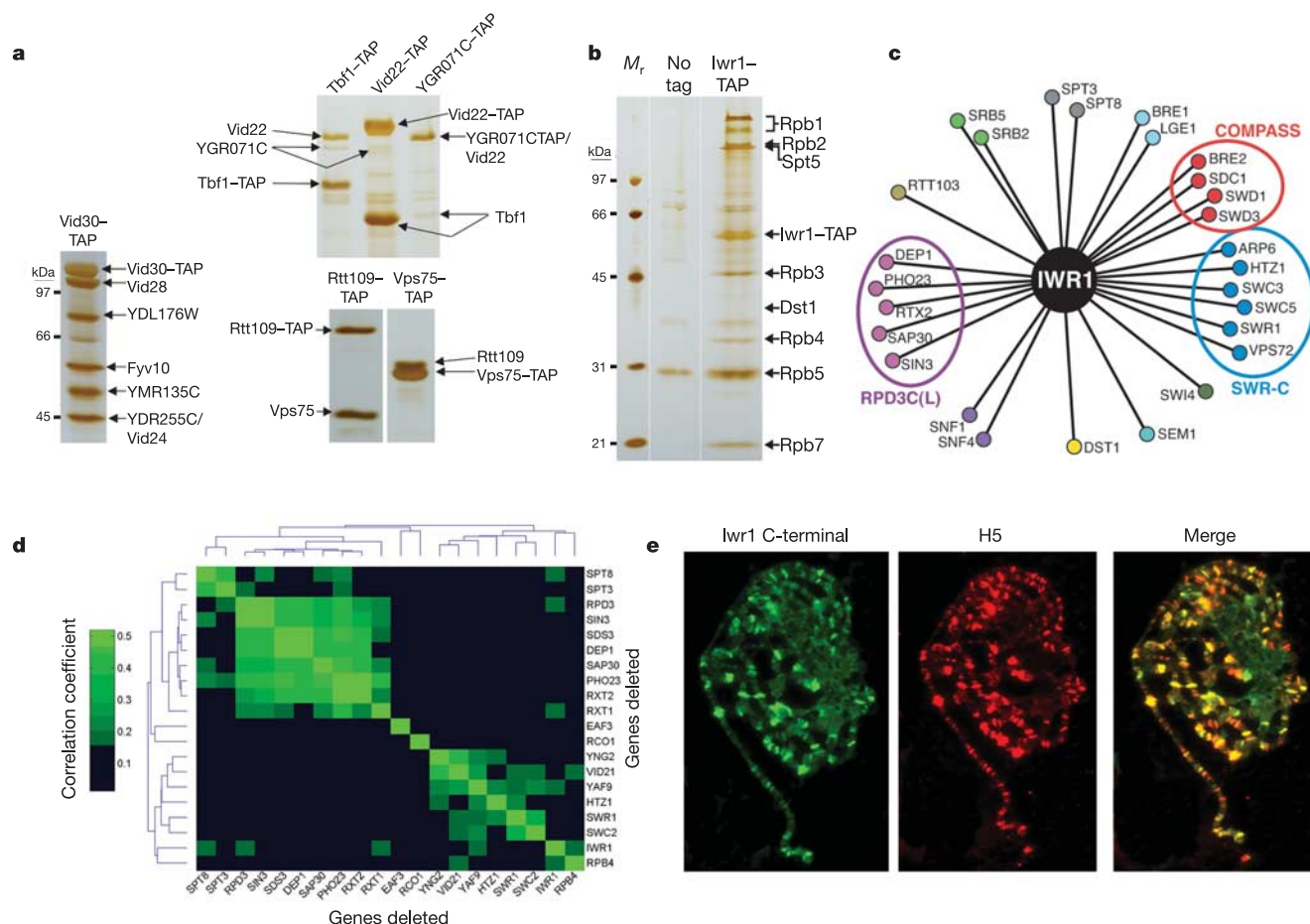


Figure 4 | Characterization of three previously unreported protein complexes and Iwr1, a novel RNAPII-interacting factor. **a**, Identification of three novel complexes by SDS-PAGE, silver staining and mass spectrometry. The same novel complex containing Vid30 was obtained after purification from strains with other tagged subunits (data not shown). **b**, Identification of Iwr1 (interacts with RNAPII). Tagging and purification of unique RNAPII subunits identified YDL115C (Iwr1) as a novel RNAPII-associated factor (Supplementary Fig. S5a). Purification of Iwr1 is shown here. **c**, Genetic interactions of Iwr1 with various transcription factors. Lines connect genes

with synthetic lethal/sick genetic interactions. **d**, Microarray analysis on the indicated deletion strains. Pearson correlation coefficients were calculated for the effects on gene expression of each deletion pair and organized by two-dimensional hierarchical clustering. **e**, Antibody generated against the amino-terminal amino acid sequence (DDDDDDDSFASADGE) of the *Drosophila* homologue of Iwr1 (CG10528) and a monoclonal antibody (H5) against RNAPII subunit Rpb1 phosphorylated on S5 of the heptapeptide repeat of its carboxy-terminal domain⁴⁸ were used for co-localization studies on polytene chromosomes as previously described⁴⁷.

high proportion of fungi-specific subunits. Previous studies have shown that highly connected proteins within a network tend to be more highly conserved^{17,34}, a consequence of either functional constraints or preferential interaction of new proteins with existing highly connected proteins²⁸. For the network as a whole, and consistent with earlier studies, Fig. 3e reveals that the frequency of ORFs with a large number (>10) of connections is proportional to the relative distance of the evolutionary group. 'Betweenness' provides a measure of how 'central' a protein is in a network, typically calculated as the fraction of shortest paths between node pairs passing through a node of interest. Figure 3e shows that highly conserved proteins tend to have higher values of betweenness. Despite these average network properties, the subunits of some complexes (for example, the kinetochore complex) display a high degree of connectedness despite restriction to hemiascomycetes. These findings suggest caution in extrapolating network properties to the properties of individual complexes. We also investigated the relationship between an ORF's essentiality and its conservation, degree of connectivity and betweenness (Supplementary Fig. S4). Consistent with previous studies^{17,35}, essential genes tend to be more highly conserved, highly connected and central to the network (as defined by betweenness), presumably reflecting their integrating role.

Examples of new protein complexes and interactions

Among the 275 complexes not in MIPS that we identified three are shown in Fig. 4a. One contains Tbf1, Vid22 and YGR071C. Tbf1 binds subtelomeric TTAGGG repeats and insulates adjacent genes from telomeric silencing^{36,37}, suggesting that this trimeric complex might be involved in this process. Consistent with this, a hypomorphic DAmP allele¹⁰ (3' untranslated region (UTR) deletion) of the essential *TBF1* gene causes a synthetic growth defect when combined with a deletion of *VID22* (data not shown), suggesting that Tbf1 and Vid22 have a common function. Vid22 and YGR071C are the only yeast proteins containing BED Zinc-finger domains, thought to mediate DNA binding or protein–protein interactions³⁸, suggesting that each uses its BED domain to interact with Tbf1 or enhance DNA binding by Tbf1. Another novel complex in Fig. 4a contains Vid30 and six other subunits (also see Fig. 3d inset). Five of its subunits (Vid30, Vid28, Vid24, Fyv10, YMR135C) have been genetically linked to proteasome-dependent, catabolite-induced degradation of fructose-1,6-bisphosphatase³⁹, suggesting that the remaining two subunits (YDL176W, YDR255C), hypothetical proteins of hitherto unknown function, are probably involved in the same process. Vid24 was reported to be in a complex with a M_r of approximately 600,000 (ref. 39), similar to the sum of the apparent M_r values of the subunits of the Vid30-containing complex. The third novel complex contains Rtt109 and Vps75. Because Vps75 is related to nucleosome assembly protein Nap1, and Rtt109 is involved in Ty transposition⁴⁰, this complex may be involved in chromatin assembly or function.

Our systematic characterization of complexes by TAP and mass spectrometry has often led to the identification of new components of established protein complexes (Fig. 3a)^{41–43}. Figure 4 highlights Iwr1 (YDL115C), which co-purifies with RNA polymerase II (RNAPII) along with general initiation factor TFIIF and transcription elongation factors Spt4/Spt5 and Dst1 (TFIIS) (Figs 4b and 3d (inset); see also Supplementary Fig. S5a). We used synthetic genetic array (SGA) technology⁹ in a quantified, high-density E-MAP format¹⁰ to systematically identify synthetic genetic interactions for *iwr1Δ* with deletions of the elongation factor gene *DST1*, the SWR complex that assembles the variant histone Htz1 into chromatin⁴⁴, an Rpd3-containing histone deacetylase complex (Rpd3(L)) that mediates promoter-specific transcriptional repression^{30,31}, the histone H3 K4 methyltransferase complex (COMPASS), the activity of which is linked to elongation by RNAPII⁴⁵, and other transcription-related genes (Fig. 4c). Moreover, DNA microarray analyses of the effects on gene expression of deletions of *IWR1* and other genes

involved in transcription by RNAPII, followed by clustering of the genes according to the similarity of their effects on gene expression, revealed that deletion of *IWR1* is most similar in its effects on mRNA levels to deletion of *RPB4* (Fig. 4d), a subunit of RNAPII with multiple roles in transcription⁴⁶. We also made use of the fact that Iwr1 is highly conserved (Supplementary Fig. S5b), with a homologue, CG10528, in *Drosophila melanogaster*. Fig. 4e shows that *Drosophila* Iwr1 partly co-localizes with phosphorylated, actively transcribing RNAPII on polytene chromosomes, suggesting that Iwr1 is an evolutionarily conserved transcription factor.

Conclusions

We have described the interactome and protein complexes underlying most of the yeast proteome. Our results comprise 7,123 protein–protein interactions for 2,708 proteins in the core data set. Greater coverage and accuracy were achieved compared with previous high-throughput studies of yeast protein–protein interactions as a consequence of four aspects of our approach: first, unlike a previous study using affinity purification and mass spectrometry¹⁴, we avoided potential artefacts caused by protein overproduction; second, we were able to ensure greater data consistency and reproducibility by systematically tagging and purifying both interacting partners for each protein–protein interaction; third, we enhanced coverage and reproducibility, especially for proteins of lower abundance, by using two independent methods of sample preparation and complementary mass spectrometry procedures for protein identification (in effect, up to four spectra were available for statistically evaluating the validity of each PPI); and finally, we used rigorous computational procedures to assign confidence values to our predictions. It is important to note, however, that our data represent a 'snapshot' of protein–protein interactions and complexes in a particular yeast strain subjected to particular growth conditions.

Both the quality of the mass spectrometry spectra used for protein identification and the approximate stoichiometry of the interacting protein partners can be evaluated by accessing our publicly available comprehensive database (<http://tap.med.utoronto.ca/>) that reports gel images, protein identifications, protein–protein interactions and supporting mass spectrometry data (Supplementary Information and Supplementary Fig. S6). Soon to be linked to our database will be thousands of sites of post-translational modification tentatively identified during our LC-MS/MS analyses (manuscript in preparation). The protein interactions and assemblies we identified provide entry points for studies on individual gene products, many of which are evolutionarily conserved, as well as 'systems biology' approaches to cell physiology in yeast and other eukaryotic organisms.

METHODS

Experimental procedures and mass spectrometry. Proteins were tagged, purified and prepared for mass spectrometry as previously described⁴³. Gel images, mass spectra and confidence scores for protein identification by mass spectrometry are found in our database (<http://tap.med.utoronto.ca/>). Confidence scores for protein identification by LC-MS/MS were calculated as described previously⁴³. After processing 72 database searches for each spectrum, a score of 1.25, corresponding to 99% confidence (A.P.T. and N.J.K., unpublished data), was used as a cut-off for protein identification by MALDI-TOF mass spectrometry. Synthetic genetic interactions and effects of deletion mutations on gene expression were identified as described previously³⁰. *Drosophila* polytene chromosomes were stained with dIwr1 anti-peptide antibody and H5 monoclonal antibody as previously described⁴⁷.

Identification of protein complexes. Details of the methods for identification of protein complexes and calculating their overlaps with various data sets are described in Supplementary Information.

Protein property analysis. We used previously published yeast protein localization data^{5,6}, and yeast protein properties were obtained from the SGD (<http://www.yeastgenome.org/>) and GO (<http://www.geneontology.org>) databases. Proteins expressed at high, medium or low levels have expression log values of >4, 3–4, or <3, respectively¹⁸.

Phylogenetic analysis. For each *S. cerevisiae* sequence a BLAST and TBLASTX

search was performed against each of the different organism data sets, including predicted ORFs from fully sequenced genomes, expressed sequence tag consensus sequences (obtained from <http://www.partigenedb.org>) and some raw genomic sequences. Using a BLAST bit score cut-off of 50, a taxonomic profile for each ORF was obtained by identifying sequences sharing significant similarity to at least one organism from each group. An ORF is said to be specific to each group only if it has a match to an organism within that group and not to any organism deemed to be more distantly related. Values of betweenness were calculated using the software Pajek (<http://vlado.fmf.uni-lj.si/pub/networks/pajek/>).

Received 20 December 2005; accepted 23 February 2006.

Published online 22 March 2006.

- Goffeau, A. *et al.* Life with 6000 genes. *Science* **274**, 546, 563–567 (1996).
- Hughes, T. R. *et al.* Functional discovery via a compendium of expression profiles. *Cell* **102**, 109–126 (2000).
- Martzen, M. R. *et al.* A biochemical genomics approach for identifying genes by the activity of their products. *Science* **286**, 1153–1155 (1999).
- Zhu, H. & Snyder, M. Protein chip technology. *Curr. Opin. Chem. Biol.* **7**, 55–63 (2003).
- Huh, W. K. *et al.* Global analysis of protein localization in budding yeast. *Nature* **425**, 686–691 (2003).
- Kumar, A. *et al.* Subcellular localization of the yeast proteome. *Genes Dev.* **16**, 707–719 (2002).
- Ross-Macdonald, P. *et al.* Large-scale analysis of the yeast genome by transposon tagging and gene disruption. *Nature* **402**, 413–418 (1999).
- Winzeler, E. A. *et al.* Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**, 901–906 (1999).
- Tong, A. H. *et al.* Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* **294**, 2364–2368 (2001).
- Schuldiner, M. *et al.* Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell* **123**, 507–519 (2005).
- Uetz, P. *et al.* A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. *Nature* **403**, 623–627 (2000).
- Ito, T. *et al.* A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl Acad. Sci. USA* **98**, 4569–4574 (2001).
- Gavin, A. C. *et al.* Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141–147 (2002).
- Ho, Y. *et al.* Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* **415**, 180–183 (2002).
- Xia, Y. *et al.* Analyzing cellular biochemistry in terms of molecular networks. *Annu. Rev. Biochem.* **73**, 1051–1087 (2004).
- von Mering, C. *et al.* Comparative assessment of large-scale data sets of protein–protein interactions. *Nature* **417**, 399–403 (2002).
- Butland, G. *et al.* Interaction network containing conserved and essential protein complexes in *Escherichia coli*. *Nature* **433**, 531–537 (2005).
- Ghaemmaghami, S. *et al.* Global analysis of protein expression in yeast. *Nature* **425**, 737–741 (2003).
- Rigaut, G. *et al.* A generic protein purification method for protein complex characterization and proteome exploration. *Nature Biotechnol.* **17**, 1030–1032 (1999).
- Link, A. J. *et al.* Direct analysis of protein complexes using mass spectrometry. *Nature Biotechnol.* **17**, 676–682 (1999).
- McCormack, A. L. *et al.* Direct analysis and identification of proteins in mixtures by LC/MS/MS and database searching at the low-femtomole level. *Anal. Chem.* **69**, 767–776 (1997).
- Krogan, N. J. *et al.* High-definition macromolecular composition of yeast RNA-processing complexes. *Mol. Cell* **13**, 225–239 (2004).
- Mewes, H. W. *et al.* MIPS: analysis and annotation of proteins from whole genomes. *Nucleic Acids Res.* **32**, D41–D44 (2004).
- Mitchell, T. *Machine Learning* (McGraw Hill, 1997).
- Wolpert, D. H. Stacked generalization. *Neural Netw.* **5**, 241–259 (1992).
- Jansen, R. & Gerstein, M. Analyzing protein function on a genomic scale: the importance of gold-standard positives and negatives for network prediction. *Curr. Opin. Microbiol.* **7**, 535–545 (2004).
- Jansen, R. *et al.* A Bayesian networks approach for predicting protein–protein interactions from genomic data. *Science* **302**, 449–453 (2003).
- Barabasi, A. L. & Albert, R. Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).
- Enright, A. J., Van Dongen, S. & Ouzounis, C. A. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* **30**, 1575–1584 (2002).
- Keogh, M. C. *et al.* Cotranscriptional Set2 methylation of Histone H3 lysine 36 recruits a repressive Rpd3 complex. *Cell* **123**, 593–605 (2005).
- Carrozza, M. J. *et al.* Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* **123**, 581–592 (2005).
- Lord, P. W., Stevens, R. D., Brass, A. & Goble, C. A. Investigating semantic similarity measures across the Gene Ontology: the relationship between sequence and annotation. *Bioinformatics* **19**, 1275–1283 (2003).
- Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
- Fraser, H. B., Wall, D. P. & Hirsh, A. E. A simple dependence between protein evolution rate and the number of protein–protein interactions. *BMC Evol. Biol.* **3**, 11 (2003).
- Joy, M. P., Brock, A., Ingber, D. E. & Huang, S. High-betweenness proteins in the yeast protein interaction network. *J. Biomed. Biotechnol.* **2005**, 96–103 (2005).
- Fourrel, G., Revardel, E., Koering, C. E. & Gilson, E. Cohabitation of insulators and silencing elements in yeast subtelomeric regions. *EMBO J.* **18**, 2522–2537 (1999).
- Brigati, C., Kurtz, S., Balderes, D., Vidali, G. & Shore, D. An essential yeast gene encoding a TTAGGG repeat-binding protein. *Mol. Cell. Biol.* **13**, 1306–1314 (1993).
- Aravind, L. The BED finger, a novel DNA-binding domain in chromatin-boundary-element-binding proteins and transposases. *Trends Biochem. Sci.* **25**, 421–423 (2000).
- Regelmann, J. *et al.* Catabolite degradation of fructose-1,6-bisphosphatase in the yeast *Saccharomyces cerevisiae*: a genome-wide screen identifies eight novel GID genes and indicates the existence of two degradation pathways. *Mol. Biol. Cell* **14**, 1652–1663 (2003).
- Scholes, D. T., Banerjee, M., Bowen, B. & Curcio, M. J. Multiple regulators of Ty1 transposition in *Saccharomyces cerevisiae* have conserved roles in genome maintenance. *Genetics* **159**, 1449–1465 (2001).
- Krogan, N. J. & Greenblatt, J. F. Characterization of a six-subunit holo-elongator complex required for the regulated expression of a group of genes in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **21**, 8203–8212 (2001).
- Krogan, N. J. *et al.* Proteasome involvement in the repair of DNA double-strand breaks. *Mol. Cell* **16**, 1027–1034 (2004).
- Krogan, N. J. *et al.* RNA polymerase II elongation factors of *Saccharomyces cerevisiae*: a targeted proteomics approach. *Mol. Cell. Biol.* **22**, 6979–6992 (2002).
- Korber, P. & Horz, W. SWRred not shaken; mixing the histones. *Cell* **117**, 5–7 (2004).
- Hampsey, M. & Reinberg, D. Tails of intrigue: phosphorylation of RNA polymerase II mediates histone methylation. *Cell* **113**, 429–432 (2003).
- Sampath, V. & Sadhale, P. Rpb4 and Rpb7: a sub-complex integral to multi-subunit RNA polymerases performs a multitude of functions. *IUBMB Life* **57**, 93–102 (2005).
- Eissenberg, J. C. *et al.* dELL is an essential RNA polymerase II elongation factor with a general role in development. *Proc. Natl Acad. Sci. USA* **99**, 9894–9899 (2002).
- Allison, L. A., Moyle, M., Shales, M. & Ingles, C. J. Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. *Cell* **42**, 599–610 (1985).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Chow, N. Mohammad, C. Chung and V. Fong for their assistance with the creation of the web resources. We are grateful to J. van Helden and S. Brohée for sharing information on their comparison of clustering methods before publication. This research was supported by grants from Genome Canada and the Ontario Genomics Institute (to J.F.G. and A.E.), the Canadian Institutes of Health Research (to A.E., N.J.K., J.F.G., S.J.W., S.P. and C.J.I.), the National Cancer Institute of Canada with funds from the Canadian Cancer Society (to J.F.G.), the Howard Hughes Medical Institute (to J.S.W. and E.O.), the McLaughlin Centre for Molecular Medicine (to S.J.W. and S.P.), the Hospital for Sick Children (to J.M.P.-A.), the National Sciences and Engineering Research Council (to N.J.K., T.R.H. and A.E.) and the National Institutes of Health (to A.S., M.G., A.P. and H.Y.).

Author Information Protein interaction information from this paper has been provided to the BioGRID database (<http://thebiogrid.org>), as well as the International Molecular Interaction Exchange consortium (IMEx, <http://imex.sf.net>) consisting of BIND, DIP, IntAct, MINT and MpaC (MIPS). Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.F.G. (jack.Greenblatt@utoronto.ca) or A.E. (Andrew.emili@utoronto.ca).

LETTERS

The evolution of galaxies from primeval irregulars to present-day ellipticals

Masao Mori^{1,2} & Masayuki Umemura³

Galaxy formation is believed to proceed in a ‘bottom up’ manner, starting with the formation of small clumps of gas and stars that then merge hierarchically into giant systems^{1,2}. The baryonic gas loses thermal energy by radiative cooling and falls towards the centres of the new galaxies, while supernovae blow gas out^{3,4}. Any realistic model therefore requires a proper treatment of these processes, but hitherto this has been far from satisfactory⁵. Here we report a simulation that follows evolution from the earliest stages of galaxy formation through the period of dynamical relaxation, at which point the resulting galaxy is in its final form. The bubble structures of gas revealed in our simulation (for times of less than 3×10^8 years) resemble closely high-redshift Lyman- α emitters^{6,7}. After 10^9 years, these bodies are dominated by stellar continuum radiation and then resemble the Lyman break galaxies^{8,9}, which are high-redshift star-forming galaxies. At this point, the abundance of elements heavier than helium (‘metallicity’) appears to be solar. After 1.3×10^{10} years, these galaxies resemble present-day ellipticals.

To explore the early evolution of galaxies, the coupling of the dynamics and the chemical evolution through star formation and supernova feedback needs to be treated properly^{10,11}. In particular, it is crucial to resolve accurately the thermalization of the kinetic energy released by multiple supernovae. We present an ultra-high-resolution ($1,024^3$ fixed cartesian grids) hydrodynamic simulation coupled with collisionless dynamics for dark matter particles and star particles, which is solved by an N -body method. The simulation pursues the early evolution ($< 2 \times 10^9$ yr) of a proto-galaxy as an assemblage of subgalactic condensations with a mass of $5.0 \times 10^9 M_\odot$ (where $M_\odot$ is the solar mass) building up a total mass of $10^{11} M_\odot$. The details of the numerical procedures are described in the Supplementary Information.

Figure 1 shows the results for the time sequence of star formation, gas dynamics and chemical enrichment. In the first 10^8 yr, stars form in high-density peaks within subgalactic condensations and the burst of star formation starts. Then, massive stars in the star forming regions explode as supernovae one after another. The gas in the vicinity of supernovae is quickly enriched with ejected metals, but a large amount of gas still retains low metal abundance. Consequently, the metallicity distribution becomes highly inhomogeneous on kiloparsec scales, where gas enriched as $-5 \leq [\text{O}/\text{H}] \leq -1$ coexists with virtually primordial gas (the oxygen abundance, $[\text{O}/\text{H}]$, is defined in Fig. 1 legend). As the density of the interstellar medium is lower in the outer regions of subgalactic condensations, the expansion of hot bubbles is accelerated there. At 3×10^8 yr, supernova-driven shocks collide with each other to generate super-bubbles of ~ 50 kpc and the surrounding high-density, cooled (10^4 K) shells. The dense shells undergo hydrodynamic instabilities induced by radiative cooling, eventually fragmenting into cold filaments and blobs. New stars are born in the enriched gas and subsequent supernovae again eject

heavy elements. The hot bubbles expand further by continual supernovae, and the shells sweep up the partially enriched ambient gas. The gas density in dense shells increases owing to the efficient radiative cooling, mainly through collisional excitation of neutral hydrogen. After 5×10^8 yr, the hot bubbles blow out into the intergalactic space. The rightmost panels of Fig. 1 show the structure at 10^9 yr. By this stage, the interstellar medium is recycled repeatedly. Eventually, some amounts of cool, dense filaments are left at the centre. But most of volume is filled with rarefied gas ($\sim 10^{-4} \text{ cm}^{-3}$) that has intermediate temperature ($10^{4.5} \leq T \text{ (K)} \leq 10^{6.5}$). At this epoch, the mixing of heavy elements is nearly completed.

Newly born stars trace the mixing history of the heavy elements well, because they inherit the metal abundance of gas. In Fig. 2, the star formation epoch is shown as a function of the oxygen abundance of newly formed stars. It is clearly seen that, before 10^8 yr, there is considerable variance in the oxygen abundance ($-5 \leq [\text{O}/\text{H}] \leq -1$), reflecting a very inhomogeneous distribution of enriched gas. After 10^8 yr, the merger of subgalactic condensations promotes the mixing of heavy elements. Finally, the almost complete recycling of interstellar matter erases the inhomogeneities of metal abundance. As a result, the oxygen abundance of stars converges to $-0.3 \leq [\text{O}/\text{H}] \leq 0.2$ with small dispersion. It is worth noting that the metal abundance is already at the level of solar abundance at 10^9 yr.

In Fig. 3, the spectral energy distribution (SED), the surface brightness distributions, and the star formation history are shown. The star formation rate increases at 5×10^7 yr, and reaches a peak of about $40 M_\odot \text{ yr}^{-1}$ around 1.5×10^8 yr. The burst of star formation continues until 3×10^8 yr. Then, the star formation activity gradually diminishes down to a few $M_\odot \text{ yr}^{-1}$ after 10^9 yr because supernova-driven winds have removed any remaining cold gas from the subgalactic fragments. As seen in the SED, at the earliest stages of less than 3×10^8 yr, the Lyman α ($\text{Ly}\alpha$) emission is conspicuous (it comes from high-density cooling shells) and its luminosity is more than $10^{43} \text{ erg s}^{-1}$. The $\text{Ly}\alpha$ luminosity perfectly matches that observed in $\text{Ly}\alpha$ emitters^{12,13} (LAEs). This result suggests that LAEs could correspond to an early supernova-dominated phase before 3×10^8 yr. Among theoretical models for LAEs^{7,14,15}, the present multiple supernova model is distinctive in having bubbly structure. In Fig. 4, the narrow-band image of extended LAE observed in ref. 12 is compared to the distribution of the $\text{Ly}\alpha$ emission of the simulated galaxy at 2×10^8 yr. We find that the physical extent of ~ 100 kpc and the bubbly structure produced by multiple supernovae are quite similar to the observed features in the $\text{Ly}\alpha$ surface brightness distribution of this LAE.

After 3×10^8 yr, the $\text{Ly}\alpha$ luminosity quickly declines to several times $10^{41} \text{ erg s}^{-1}$, as the emission from cooling gas decreases immediately owing to the leak of explosion energy through the blowouts of super-bubbles. Then, the SED becomes dominated by

¹Department of Physics and Astronomy, University of California, Los Angeles, California 90095-1547, USA. ²Institute of Natural Sciences, Senshu University, Kawasaki, Kanagawa 214-8580, Japan. ³Center for Computational Sciences, University of Tsukuba, Tsukuba, Ibaraki 305-8577, Japan.

stellar continuum emission. The galaxy in this phase features diffuse, asymmetric structures, and outflows of $100\text{--}500\text{ km s}^{-1}$. The total mass of long-lived stars is $9.3 \times 10^9 M_\odot$, and a mass of $1.5 \times 10^9 M_\odot$ is involved in the outflows at redshift $z = 3$. These features look quite similar to those observed for Lyman break galaxies^{16,17} (LBGs). The low-ionization interstellar absorption lines observed in LBGs are blueshifted by hundreds of km s^{-1} relative to systemic velocities and Ly α lines are redshifted to the same degree. Furthermore, the strong metal absorption lines observed in the spectra of LBGs indicate that their star formation events must have been preceded by an earlier starburst. The excess of absorption-line systems with large C IV column density in spectra of background quasars near LBGs is interpreted as further evidence for chemical enrichment of the intergalactic medium due to the supernova-driven outflows. Recently, the X-ray luminosity^{18,19} at $2.0\text{--}8.0\text{ keV}$ for LBGs has been

found to be $\sim 10^{41}\text{ erg s}^{-1}$. In the present simulation, the X-ray luminosity at the same energy range changes from $10^{42}\text{ erg s}^{-1}$ at $3 \times 10^8\text{ yr}$ to $\sim 10^{41}\text{ erg s}^{-1}$ around 10^9 yr . The LBG metallicity appears to be the solar value for massive systems²⁰. In the light of such properties, the simulated post-starburst galaxy with an age of 10^9 yr can correspond to LBGs. Thus, it is implied that LBGs are the next phase of LAEs.

The long-term dynamical evolution of the model galaxy was studied with an N -body simulation containing one million particles. We found that the assembly of subcondensations and the virialization of the total system are almost completed in $3 \times 10^9\text{ yr}$, so that the system achieves a quasi-equilibrium state. The resultant stellar system forms a virialized, spheroidal system. Figure 3b shows the projected surface brightness distributions in the U, B, V and K bands at $1.3 \times 10^{10}\text{ yr}$ ($z = 0$) assuming passive evolution (no further star

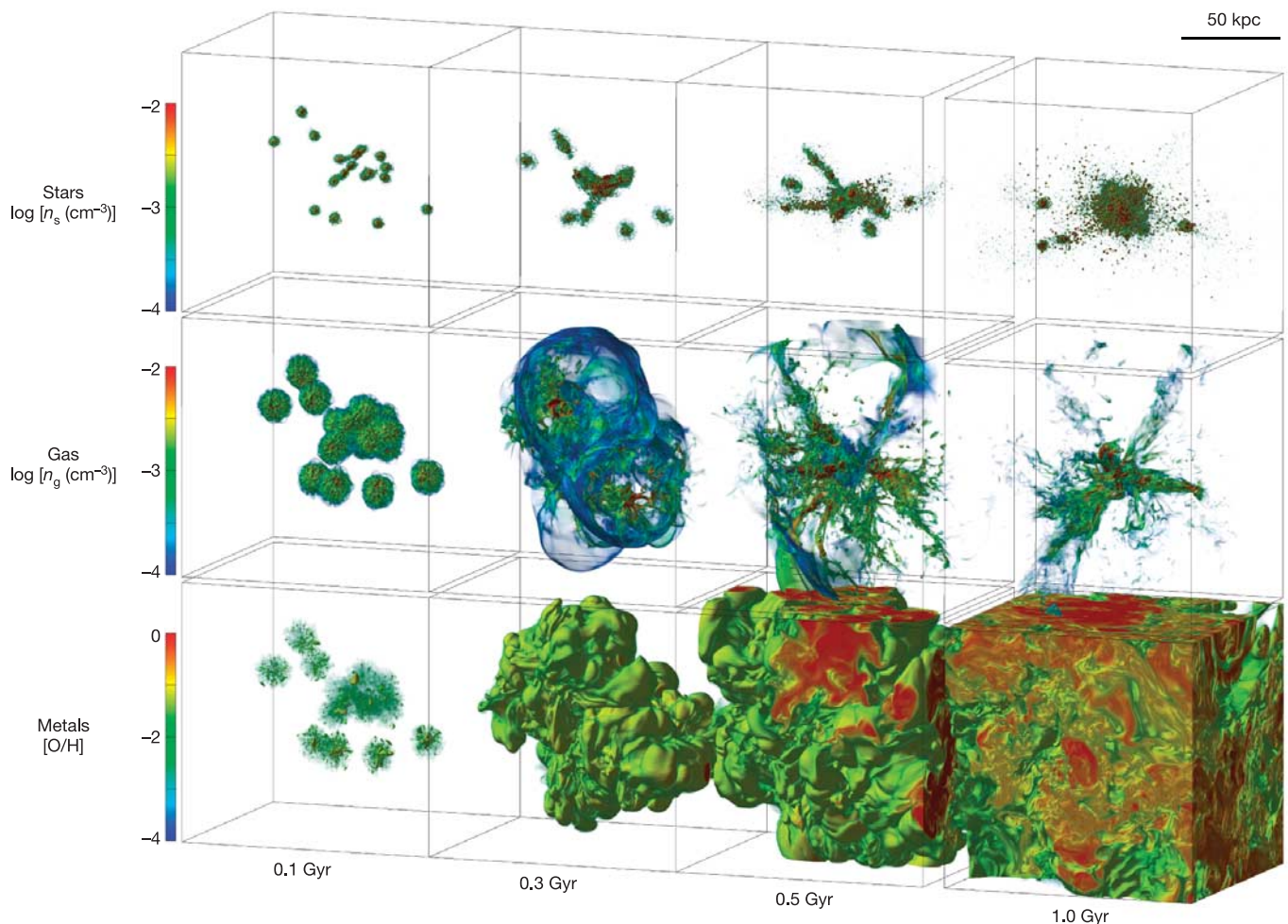


Figure 1 | Simulation of the first 1 Gyr of a proto-galaxy with total mass $10^{11} M_\odot$ (1 Gyr = 10^9 yr , $M_\odot$ indicates solar mass). The overdensity region of this mass-scale decouples from the cosmic expansion at redshift $z = 7.8$ at a radius of 53.7 kpc (where 1 kpc = 3,260 light yr), where the initial conditions are set up. The mass of gaseous matter is $1.3 \times 10^{10} M_\odot$ initially. The angular momentum is provided by a spin parameter of $\lambda = 0.05$ (ref. 26). Here, we assume the Λ CDM cosmology with $\Omega_M = 0.3$, $\Omega_b = 0.04$ and $\Omega_\Lambda = 0.7$, where Ω_M is the matter density, Ω_b the baryon density, and Ω_Λ the cosmological constant. The Hubble constant is assumed to be $H_0 = 70\text{ km s}^{-1}\text{ Mpc}^{-1}$. The density profiles in subgalactic dark haloes are given by the Navarro–Frenk–White profile²⁷ and these condensations are distributed randomly within the galaxy-scale overdensity. Radiative cooling for the gaseous component is calculated using the cooling function for an optically thin, collisionally ionized gas²⁸. Stars are assumed to form in

gravitationally unstable cooled regions with Salpeter's initial mass function²⁹, at a rate that is inversely proportional to the local free-fall time^{10,22}. Stars more massive than $8 M_\odot$ explode as type II supernovae with an explosion energy of 10^{51} erg , and eject synthesized heavy elements. The evolution is shown by the spatial distributions of the stellar density (n_s ; top row), the gas density (n_g ; middle row) and the oxygen abundance ($[\text{O}/\text{H}]$, bottom row). Here $[\text{O}/\text{H}] = \log_{10}(N_{\text{O}}/N_{\text{H}}) - \log_{10}(N_{\text{O}}/N_{\text{H}})_\odot$ for gas, where N_{O} and N_{H} are the number densities of oxygen and hydrogen, respectively. Each simulation box has a physical size of 134 kpc and the spatial resolution is 0.131 kpc. This is comparable with the typical size of super-bubbles observed in the local Universe. Both the number density of the stellar component and that of the gas component range from 10^{-4} cm^{-3} to 10^{-2} cm^{-3} , and the gas metallicity ranges from -4 to 0 .

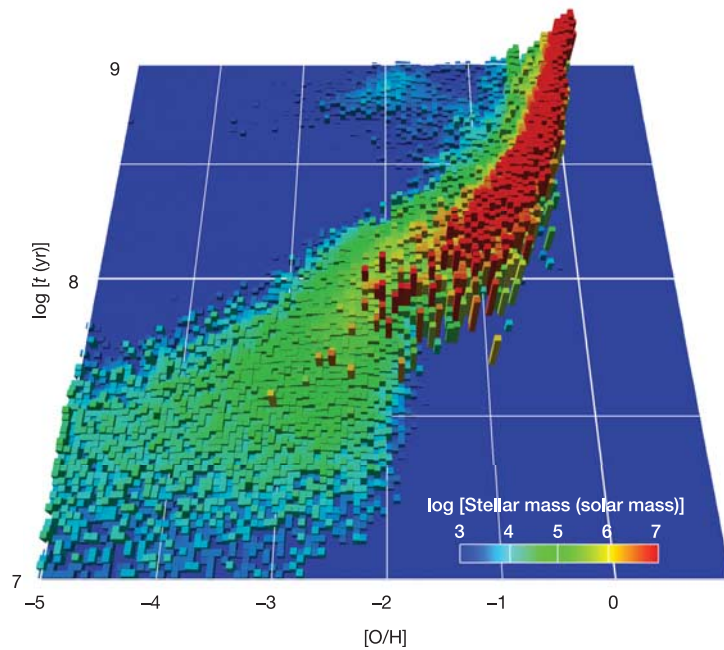


Figure 2 | The formation epochs (t) of stars as a function of stellar oxygen abundance, $[O/H]$. The colour-coded histogram shows the stellar mass on logarithmic scales.

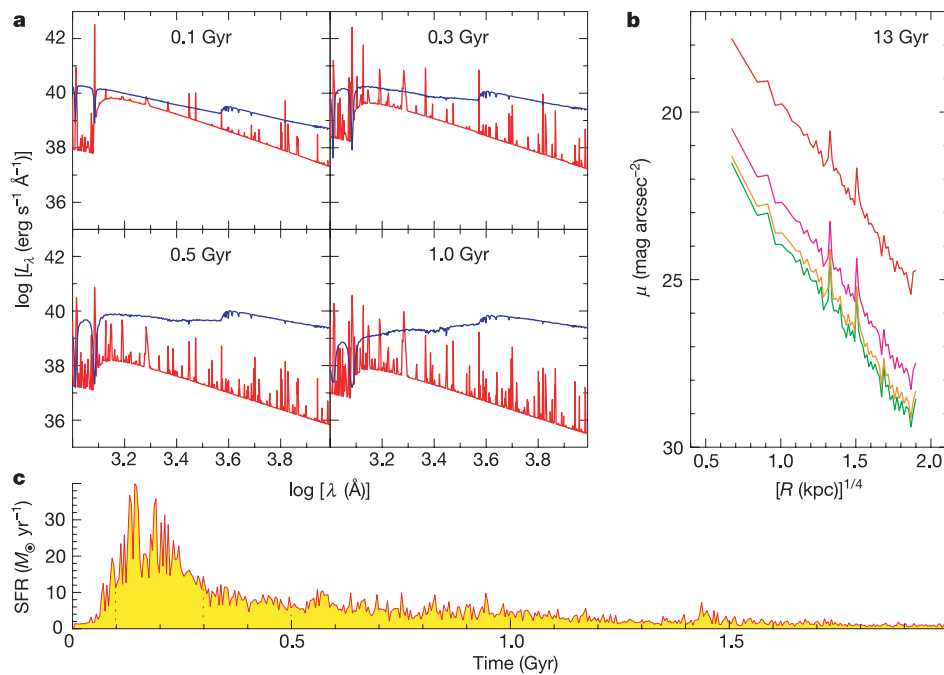


Figure 3 | Emissions and star formation history. **a**, Predicted spectral energy distribution (SED) of the emission from the simulated galaxy. The emission properties of the gas components are calculated for an optically thin, collisionally ionized gas using the MAPPINGIII code²⁸ (red lines), and those of the stellar components are calculated using the evolutionary stellar population synthesis code PÉGASE³⁰ (blue lines). In practice, to obtain the SED, we sum up the SED of each grid point for the gas components and each star particle for the stellar components. Here L_λ is the luminosity per unit

wavelength λ . The absolute luminosities of Ly α line emission, where the wavelength is 1,216 Å, are $2.0 \times 10^{43} \text{ erg s}^{-1}$, $1.6 \times 10^{43} \text{ erg s}^{-1}$, $4.6 \times 10^{41} \text{ erg s}^{-1}$ and $2.3 \times 10^{41} \text{ erg s}^{-1}$ at an elapsed time of 0.1 Gyr, 0.3 Gyr, 0.5 Gyr and 1 Gyr, respectively. **b**, Projected distribution of surface brightness at 13 Gyr for our simulation run. Solid lines from bottom to top are the surface brightness μ in the U, B, V and K bands, respectively, plotted against the quartic root of the radius R . **c**, Star formation rate as a function of time.

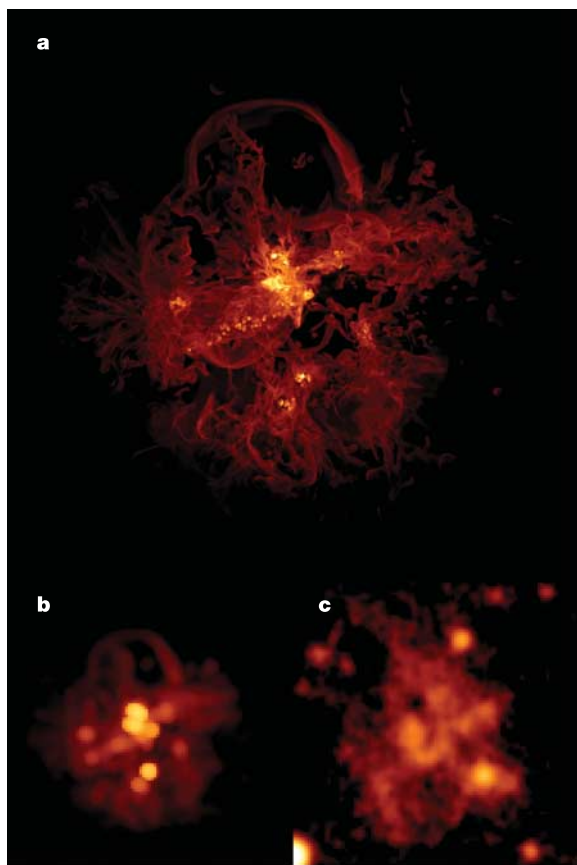


Figure 4 | Comparison of the simulation and observation. All panels are 154 kpc square. **a**, The projected distribution of Ly α emission from the gas component for the simulated galaxy at 2.0×10^8 yr. **b**, The simulation result smoothed with a gaussian kernel with a full-width at half-maximum of 7.6 kpc, which corresponds to $1.0''$ at redshift $z = 3.09$, the same resolution (100 pixels $\times$ 100 pixels) as the observation¹². **c**, Narrow-band image of the extended Ly α emitter 'LAB1' taken with the Subaru Telescope in the SSA22 field at redshift $z = 3.09$ (ref. 12).

formation). They have a large central concentration that accords well with de Vaucouleurs' $r^{1/4}$ profile²¹, which is commonly found in nearby elliptical galaxies^{11,22}. The resultant absolute magnitudes in the blue band (B) and the visual band (V) are $M_B = -17.2$ mag and $M_V = -18.0$ mag, respectively. The colours $U - V = 1.15$ and $V - K = 2.85$ are consistent with the colour-magnitude relation of elliptical galaxies in the Coma cluster of galaxies²³. Furthermore, the combination of the surface brightness, the effective radius $r_e = 3.97$ kpc, and the central velocity dispersion $\sigma_0 = 133 \text{ km s}^{-1}$ is on the fundamental plane of elliptical galaxies within their scatters. (The fundamental plane is the relationship among these three parameters derived for nearby elliptical galaxies^{24,25}.) Thus, it is suggested that LBGs evolve into elliptical galaxies through purely collisionless dynamical evolution.

Received 1 September; accepted 15 December 2005.

- Blumenthal, G. R., Faber, S. M., Primack, J. R. & Rees, M. J. Formation of galaxies and large-scale structure with cold dark matter. *Nature* **311**, 517–525 (1984).
- Springel, V. *et al.* Simulations of the formation, evolution and clustering of galaxies and quasars. *Nature* **435**, 629–636 (2005).
- Mathews, W. G. & Baker, J. C. Galactic winds. *Astrophys. J.* **170**, 241–259 (1971).
- Mori, M., Ferrara, A. & Madau, P. Early metal enrichment by pregalactic

outflows. II. Three-dimensional simulations of blow-away. *Astrophys. J.* **571**, 40–55 (2002).

- Mori, M., Umemura, M. & Ferrara, A. The nature of Ly α blobs: supernova-dominated primordial galaxies. *Astrophys. J.* **613**, L97–L100 (2004).
- Dey, A. *et al.* A galaxy at $z = 5.34$. *Astrophys. J.* **498**, L93–L97 (1998).
- Taniguchi, Y. *et al.* Lyman α emitters beyond redshift 5: The dawn of galaxy formation. *J. Korean Astron. Soc.* **36**, 123–144 (2003).
- Steidel, C. C. *et al.* Spectroscopic confirmation of a population of normal star-forming galaxies at redshifts $z > 3$. *Astrophys. J.* **462**, L17–L21 (1996).
- Giallisco, M. Lyman-break galaxies. *Annu. Rev. Astron. Astrophys.* **40**, 579–641 (2002).
- Mori, M., Yoshii, Y., Tsujimoto, T. & Nomoto, K. The evolution of dwarf galaxies with star formation in an outward-propagating supershell. *Astrophys. J.* **478**, L21–L24 (1997).
- Sommer-Larsen, J., Götz, M. & Portinari, L. Galaxy formation: Cold dark matter, feedback and the Hubble sequence. *Astrophys. J.* **596**, 47–66 (2003).
- Matsuda, Y. *et al.* A SUBARU search for Ly α blobs in and around the protocluster region at redshift $z = 3.1$. *Astron. J.* **128**, 569–584 (2004).
- Taniguchi, Y. *et al.* The SUBARU Deep Field Project: Lyman α emitters at a redshift of 6.6. *Publ. Astron. Soc. Jpn* **57**, 165–182 (2005).
- Haiman, Z., Spaans, M. & Quataert, E. Ly α cooling radiation from high-redshift halos. *Astrophys. J.* **537**, L5–L8 (2000).
- Fardal, M. A. *et al.* Cooling radiation and the Ly α luminosity of forming galaxies. *Astrophys. J.* **562**, 605–617 (2001).
- Pettini, M. *et al.* The rest-frame optical spectra of Lyman break galaxies: star formation, extinction, abundances, and kinematics. *Astrophys. J.* **554**, 981–1000 (2001).
- Adelberger, K. L., Steidel, C. C., Shapley, A. E. & Pettini, M. Galaxies and intergalactic matter at redshift $z = 3$: Overview. *Astrophys. J.* **584**, 45–75 (2003).
- Brandt, W. N. *et al.* The Chandra Deep Field-North Survey. VII. X-ray emission from Lyman break galaxies. *Astrophys. J.* **558**, L5–L9 (2001).
- Nandra, K. *et al.* X-ray properties of Lyman break galaxies in the Hubble Deep Field-North region. *Astrophys. J.* **576**, 625–639 (2002).
- Shapley, A. E. *et al.* Evidence for solar metallicities in massive star-forming galaxies at $z \sim 2$. *Astrophys. J.* **612**, 108–121 (2004).
- de Vaucouleurs, G. Recherches sur les nébuleuses extragalactiques. *Ann. Astrophys.* **11**, 247–287 (1948).
- Mori, M., Yoshii, Y. & Nomoto, K. Dissipative process as a mechanism of differentiating internal structures between dwarf and normal elliptical galaxies in a cold dark matter universe. *Astrophys. J.* **511**, 585–594 (1999).
- Bower, R. G., Lucey, J. R. & Ellis, R. S. Precision photometry of early type galaxies in the Coma and Virgo clusters - a test of the universality of the colour / magnitude relation—Part two - analysis. *Mon. Not. R. Astron. Soc.* **254**, 601–613 (1992).
- Djorgovski, S. & Davis, M. Fundamental properties of elliptical galaxies. *Astrophys. J.* **313**, 59–68 (1987).
- Burstein, D., Bender, R., Faber, S. & Nolthenius, R. Global relationships among the physical properties of stellar systems. *Astron. J.* **114**, 1365–1392 (1997).
- Barnes, J. & Efstathiou, G. Angular momentum from tidal torques. *Astrophys. J.* **319**, 575–600 (1987).
- Navarro, J. F., Frenk, C. S. & White, S. D. M. A universal density profile from hierarchical clustering. *Astrophys. J.* **490**, 493–508 (1997).
- Sutherland, R. S. & Dopita, M. A. Cooling functions for low-density astrophysical plasmas. *Astrophys. J.* **88** (Suppl.), 253–327 (1993).
- Salpeter, E. E. The luminosity function and stellar evolution. *Astrophys. J.* **121**, 161–167 (1955).
- Fioc, M. & Rocca-Volmerange, B. PEGASE: a UV to NIR spectral evolution model of galaxies. Application to the calibration of bright galaxy counts. *Astron. Astrophys.* **326**, 950–962 (1997).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank Y. Matsuda and his collaborators for use of observational data obtained by the Subaru Telescope, and are grateful to M. Rich, M. Malkan, I. Saviane, Y. Yoshii and R. Ellis for suggestions. M.M. thanks the University of California Los Angeles for hospitality, and acknowledges the support of the Japan Society for the Promotion of Science and of the Promotion and Mutual Aid Corporation for Private Schools of Japan. M.U. acknowledges the support of the Ministry of Education, Culture, Sports, Science, and Technology of Japan. The simulations were performed with the Earth Simulator at JAMSTEC, the SPACE at Senshu University, and the computational facilities including CP-PACS at CCS in the University of Tsukuba.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.M. (mmori@isc.senshu-u.ac.jp).

LETTERS

100-metre-diameter moonlets in Saturn's A ring from observations of 'propeller' structures

Matthew S. Tiscareno¹, Joseph A. Burns^{1,2}, Matthew M. Hedman¹, Carolyn C. Porco³, John W. Weiss³, Luke Dones⁴, Derek C. Richardson⁵ & Carl D. Murray⁶

Saturn's main rings are composed predominantly of water-ice particles ranging between about 1 centimetre and 10 metres in radius. Above this size range, the number of particles drops sharply, according to the interpretation of spacecraft¹ and stellar² occultations. Other than the gap moons Pan and Daphnis (the provisional name of S/2005 S1), which have sizes of several kilometres, no individual bodies in the rings have been directly observed, and the population of ring particles larger than ten metres has been essentially unknown. Here we report the observation of four longitudinal double-streaks in an otherwise bland part of the mid-A ring. We infer that these 'propeller'-shaped perturbations^{3–5} arise from the effects of embedded moonlets approximately 40 to 120 m in diameter. Direct observation of this phenomenon validates models of proto-planetary disks in which similar processes are posited^{4,6}. A population of moonlets, as implied by the size distribution that we find, could help explain gaps in the more tenuous regions of the Cassini division and the C ring⁷. The existence of such large embedded moonlets is most naturally compatible with a ring originating in the break-up of a larger body^{8–11}, but accretion from a circumplanetary disk¹² is also plausible if subsequent growth onto large particles occurs after the primary accretion phase has concluded^{13,14}.

Four examples of a unique structure previously unseen in the rings were found in two images (Fig. 1) taken by the Imaging Science Subsystem (ISS) of the Cassini spacecraft. Each of these features is a symmetric double-streak, the individual lobes of which lie in the longitudinal (horizontal) direction, with a radial (vertical) offset between them. In each case, the lobe that is radially closer to Saturn also extends in the longitudinally leading direction (that is, in the direction of orbital motion). Supplementary Figs S1 and S2 display the full images, and their placement within the ring system is given in Supplementary Fig. S3.

'Propeller'-shaped structures, very similar to those visible here, have been predicted analytically^{3,4} and simulated numerically⁵. Such disturbances^{15,16} are produced when background ring particles are carried by the Keplerian shear flow past a more massive compatriot. Moonlets larger than a few kilometres have been predicted^{7,17} to clear gaps that extend the full circumference of the rings, just as Pan and Daphnis are seen to do. In contrast, the perturbations introduced by smaller moonlets are washed out as diffusive and viscous effects quickly fill in the disturbed region. For perturbing embedded moonlets of intermediate size—tens to hundreds of metres in radius—the resulting disturbance has two interwoven components: an S-shaped gap (with reduced, but non-zero, density), flanked by density enhancements generated similarly to the 'moonlet wakes' present on either side of the Encke and Keeler gaps^{18–20}.

The observed 'propeller' features are two to three times brighter

than the background ring (see Fig. 2). Because the images under discussion show the unlit side of the rings, bright features may in principle be either more or less dense than the surrounding ring material (an entirely evacuated gap in the rings would scatter no light, and a completely opaque ring would transmit no light). Given previous measurements of the background optical depth in the mid-A ring^{21,22}, the near-nadir observing geometry for these images, and standard photometric models^{23,24}, we expect to find that bright features correspond to density enhancements. However, these models predict significantly lower contrast between dense and background regions than is observed. We note that Voyager images of the A ring similarly exhibited high contrast that could not be explained by standard photometric models²⁴. Differences in ring thickness between the 'propeller' structure and the background ring may affect the photometric behaviour in unknown ways (especially considering the unique viewing geometry of these images). Furthermore, the presence of self-gravity wakes^{15,25–27} pervading the surrounding ring should lower the background ring's optical depth, and hence brightness, from the standard model predictions. The absence of wakes in the perturbed 'propeller' regions may explain the increase in contrast.

Figure 2 plots the locations of the brightness enhancements seen in Fig. 1, from which we measure the mean radial position of each lobe and then the radial offset Δr . Although the perturber's radius is directly proportional to the radial separation between the gaps⁵, such a relationship is less clear for the related density enhancements. Thus, although the radial offsets are measured with $\sim 10\%$ uncertainty, model dependence dominates the uncertainty in the inferred moonlet sizes. Our observations are consistent with moonlets of the order of 20–60 m in radius embedded in the A ring, with the larger sizes being inferred when the bright features are interpreted as gaps.

Figure 2 also shows longitudinal scans along the features, in which pixel brightnesses at the core of each double-streak are radially binned and summed. The profile has a steeper slope on the side facing the perturber, just as numerical simulations produce. The full longitudinal extent of the 'propeller' features is ~ 3 km. Radial scans across the features were also computed for these images (see Supplementary Fig. S4), and show symmetrical gaussian shapes with widths similar to the radial offsets.

The rings' dynamical viscosity can in principle be derived from the length of the 'propeller' features in the longitudinal direction (effectively, the time it takes for diffusive processes to 'fill in' the disturbance created by the moonlet). The viscosity is significantly influenced by self-gravity wakes²⁸, with a theoretically expected value of $\nu \approx 90 \text{ cm}^2 \text{ s}^{-1}$ for this location in the rings ($a = 130,000 \text{ km}$). However, the uncertain photometry (see above) hampers our efforts at obtaining a meaningful viscosity measurement in multiple ways. Not only does the bright/dark ambiguity leave the moonlet's size

¹Department of Astronomy, Cornell University, ²Department of Theoretical and Applied Mechanics, Cornell University, Ithaca, New York 14853, USA. ³CICLOPS, Space Science Institute, 4750 Walnut Street, Boulder, Colorado 80301, USA. ⁴Southwest Research Institute, 1050 Walnut Street, Boulder, Colorado 80302, USA. ⁵Department of Astronomy, University of Maryland, College Park, Maryland 20742, USA. ⁶Astronomy Unit, Queen Mary, University of London, Mile End Road, E1 4NS, UK.

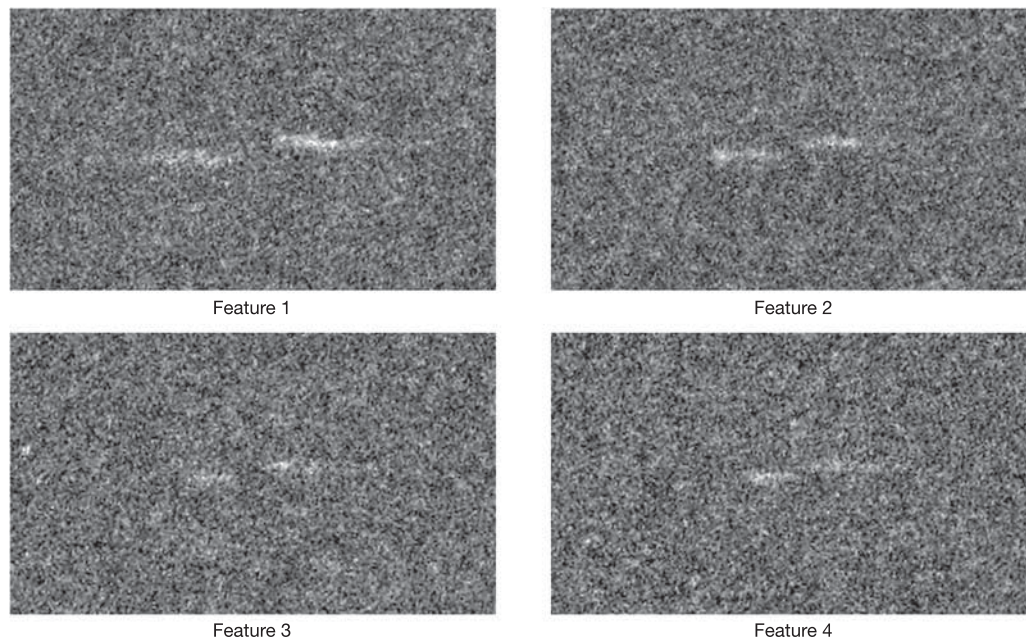


Figure 1 | Four longitudinally aligned double-streaks observed in a bland region of Saturn's A ring by the Cassini ISS camera. These are interpreted to be regions perturbed by unseen embedded moonlets located centrally between the streaks. The images have been cropped and reprojected, so that orbital motion is to the right, and Saturn's direction (radially inward) is up. In each of the four cases, the upper right-hand streak is closer to Saturn and orbitally leads the unseen moon. Cassini images N1467347210 (feature 1) and N1467347249 (features 2–4), seen in their entirety as Supplementary Figs S1 and S2, are the highest-resolution ring images yet obtained by

Cassini, and were taken during the spacecraft's insertion into Saturn orbit¹⁹ on 1 July 2004. The images were calibrated using standard techniques³⁰ to convert discrete pixel data numbers to units of brightness divided by the solar flux (I/F). Residual horizontal banding (on the level of a few data numbers) was removed by horizontally averaging pixels away from the features of interest. The nominal image resolution is 52 m per pixel, and smearing due to keplerian motion of ring particles amounts to less than three pixels.

uncertain, but it is similarly difficult to calibrate absolutely the optical depth at which our data fall below the noise level. With plausible assumptions on these matters, our observations imply that ν ranges from $\sim 0.1 \text{ cm}^2 \text{ s}^{-1}$ to $\sim 700 \text{ cm}^2 \text{ s}^{-1}$.

Since four objects were found in a pair of images covering $2,800 \text{ km}^2$ apiece, we estimate the surface number density of moonlets

approximately 50 m in radius in weakly perturbed portions of the A ring to be $7 \times 10^{-4} \text{ km}^{-2}$. The total surface area of the A ring is $\sim 1.2 \times 10^{10} \text{ km}^2$, giving an estimated population (perhaps a primordial population, later altered in the more perturbed regions) of some 10^7 moonlets of this size. This calculation does not include two other images of similar resolution taken in the same sequence, in

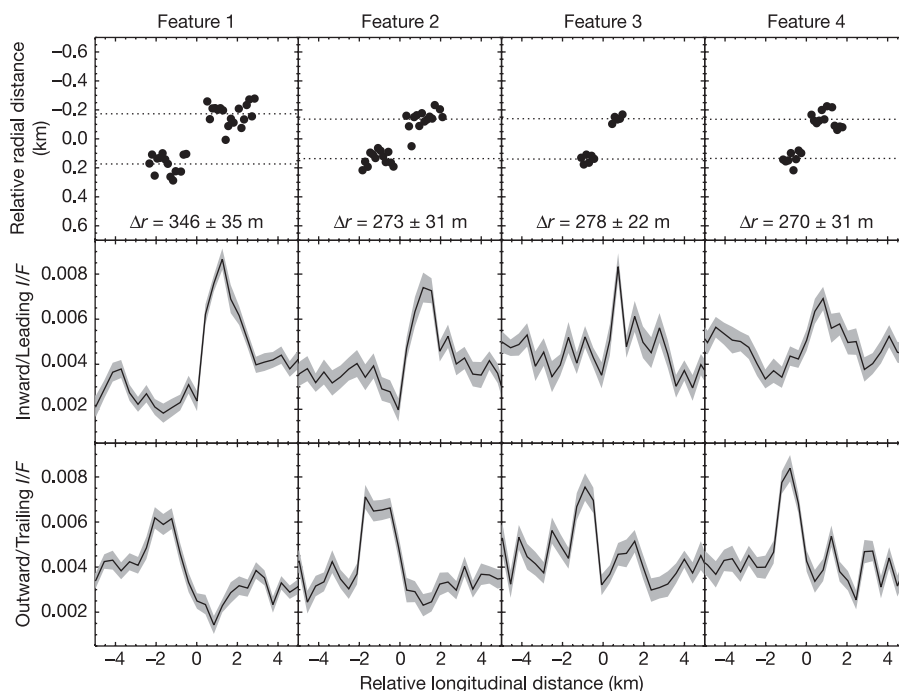


Figure 2 | Peak locations and longitudinal scans of the four double-streaks in Fig. 1. Top row, the radial locations of the brightest part of each feature, as a function of longitudinal distance relative to the inferred moon, are found by a gaussian fit to the total brightness. To increase the quality of each fit, resolution was lowered to 260 m in the longitudinal direction. We discarded four of the 90 data points because the gaussian's centre fell on a point of noise rather than the point of interest. Note the radial offset between the peaks of the outer and inner lobes. Dotted lines show the mean values for each lobe, and the radial offset between them, Δr , is given in each panel. Middle row, longitudinal scans of the brightness I/F along the radius of the inner lobe of each double-streak. The grey regions surrounding the solid lines denote the standard deviation of the mean value of the pixels in each bin. Bottom row, longitudinal scans of I/F along the radius of the outer lobe of each double-streak. We note that the longitudinal profiles are generally steeper on the side facing the moonlet.

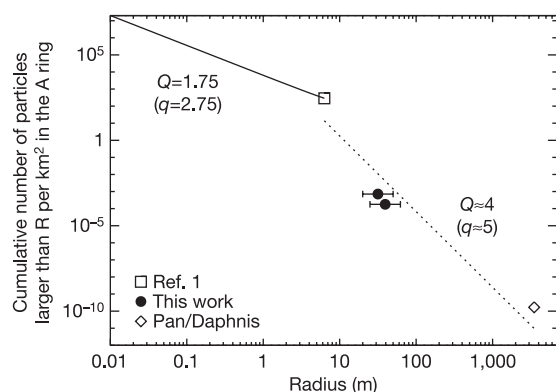


Figure 3 | Cumulative size distribution for particles in the A ring. The solid line and open square are calculated from Voyager radio occultations¹; solid circles denote the moonlets announced in this work, with the plotted error bars indicating the model-dependent uncertainties (the range of possible radii) in their sizes; the open diamond indicates the two known ring moons Pan and Daphnis. The fitted cumulative power law for particles over ten metres in size (dotted line) has an index $Q = 4 \pm 1$. This 1σ uncertainty of ± 1 comes from linear regression using the moonlet size that gives the highest residuals, thus accounting for the model-dependent uncertainties. For a differential power law, such as is discussed in the main text, this corresponds to $q = Q + 1 = 5$.

which no features of this kind were found; we attribute this lack to the stronger density waves¹³ present in those regions, which probably modify the moonlet population.

These findings allow us to extend previous estimates of the size distribution of particles in Saturn's rings. Interpretations of occultations of Voyager radio signals¹ and stars² have inferred a differential power-law distribution, $dn(R) \approx R^{-q}$ (where dn is the number of particles per unit area with radius R in the differential bin dR), with $2.7 < q < 3$ for centimetre-size to metre-size particles^{1,29}. For larger particles, however, the distribution falls quite steeply. Figure 3 shows that the present results provide a 'missing link' between the largest particles observable by occultations ($r \approx 10$ m) and the two ring moons Pan and Daphnis ($r > \sim 3.5$ km). This analysis allows us to estimate a differential power-law index $q = 5 \pm 1$ over the range $10 \text{ m} < r < 3 \text{ km}$.

The lack of similar features caused by even smaller moonlets can be attributed to the insignificant amplitudes expected in their density modulations, making them difficult to discern in these noisy images despite nominally sufficient spatial resolution. 'Propellers' too tiny to be resolved in an image would create an asymmetric profile in the noise; a preliminary search for such a profile has been unsuccessful. The current non-detection of larger moonlets ($r > \sim 100$ m) may be attributed to their rarity, as implied by the steep power-law size distribution; such features will be sought in planned lower-resolution images.

The present discovery indicates that the moons Pan and Daphnis are not isolated anomalies; rather, they are the endmembers in a continuous population of ring particles and embedded moonlets with a steep power-law size distribution. The largest bodies expected from direct accretion are on the order of the Toomre scale length, $L \approx 10$ m for the A ring¹³, though subsequent accretion of ring particles may produce larger sizes^{14,27}, whereas particles up to 5 km in radius will result from the break-up of a larger moon⁸. Thus, a population of embedded moonlets 100 m in diameter will place an important constraint on the origin of Saturn's rings.

Received 9 November 2005; accepted 12 January 2006.

1. Zebker, H. A., Marouf, E. A. & Tyler, G. L. Saturn's rings: Particle size distributions for thin layer models. *Icarus* **64**, 531–548 (1985).
2. French, R. G. & Nicholson, P. D. Saturn's rings II. Particle sizes inferred from stellar occultation data. *Icarus* **145**, 502–523 (2000).
3. Spahn, F. & Sremcevic, M. Density patterns induced by small moonlets in Saturn's rings? *Astron. Astrophys.* **358**, 368–372 (2000).
4. Sremcevic, M., Spahn, F. & Duschl, W. J. Density structures in perturbed thin cold discs. *Mon. Not. R. Astron. Soc.* **337**, 1139–1152 (2002).
5. Seiss, M., Spahn, F., Sremcevic, M. & Salo, H. Structures induced by small moonlets in Saturn's rings: Implications for the Cassini Mission. *Geophys. Res. Lett.* **32**, L11205, doi:10.1029/2005GL022506 (2005).
6. Bryden, G., Chen, X., Lin, D. N. C., Nelson, R. P. & Papaloizou, J. C. B. Tidally induced gap formation in protostellar disks: gap clearing and suppression of protoplanetary growth. *Astrophys. J.* **514**, 344–367 (1999).
7. Lissauer, J. J., Shu, F. H. & Cuzzi, J. N. Moonlets in Saturn's rings? *Nature* **292**, 707–711 (1981).
8. Harris, A. W. in *Planetary Rings* (eds Greenberg, R. & Brahic, A.) 641–659 (Univ. Arizona Press, Tucson, 1984).
9. Dones, L. A recent cometary origin for Saturn's rings? *Icarus* **92**, 194–203 (1991).
10. Smith, B. A. et al. A new look at the Saturn system—the Voyager 2 images. *Science* **215**, 504–537 (1982).
11. Colwell, J. E. The disruption of planetary satellites and the creation of planetary rings. *Planet. Space Sci.* **42**, 1139–1149 (1994).
12. Pollack, J. B. The rings of Saturn. *Space Sci. Rev.* **18**, 3–93 (1975).
13. Shu, F. H. in *Planetary Rings* (eds Greenberg, R. & Brahic, A.) 513–561 (Univ. Arizona Press, Tucson, 1984).
14. Weidenschilling, S. J., Chapman, C. R., Davis, D. & Greenberg, R. in *Planetary Rings* (eds Greenberg, R. & Brahic, A.) 367–415 (Univ. Arizona Press, Tucson, 1984).
15. Julian, W. H. & Toomre, A. Non-axisymmetric responses of differentially rotating disks of stars. *Astrophys. J.* **146**, 810–827 (1966).
16. Murray, C. D. & Dermott, S. F. *Solar System Dynamics* (Cambridge Univ. Press, 1999).
17. Hénon, M. A simple model of Saturn's rings. *Nature* **293**, 33–35 (1981).
18. Showalter, M. R., Cuzzi, J. N., Marouf, E. A. & Esposito, L. W. Satellite 'wakes' and the orbit of the Encke Gap moonlet. *Icarus* **66**, 297–323 (1986).
19. Porco, C. C. et al. Cassini Imaging Science: initial results on Saturn's rings and small satellites. *Science* **307**, 1226–1236 (2005).
20. Lewis, M. C. & Stewart, G. R. Expectations for Cassini observations of ring material with nearby moons. *Icarus* **178**, 124–143 (2005).
21. Esposito, L. A., O'Callahan, M. & West, R. A. The structure of Saturn's rings: Implications from the Voyager stellar occultation. *Icarus* **56**, 439–452 (1983).
22. Nicholson, P. D. et al. Saturn's rings I. Optical depth profiles from the 28 Sgr occultation. *Icarus* **145**, 474–501 (2000).
23. Cuzzi, J. N. et al. in *Planetary Rings* (eds Greenberg, R. & Brahic, A.) 73–199 (Univ. Arizona Press, Tucson, 1984).
24. Dones, L., Cuzzi, J. N. & Showalter, M. R. Voyager photometry of Saturn's A ring. *Icarus* **105**, 184–215 (1993).
25. Dones, L. & Porco, C. C. Spiral density wakes in Saturn's A Ring? *Bull. Am. Astron. Soc.* **21**, 929 (1989).
26. Salo, H. Simulations of dense planetary rings III. Self-gravitating identical particles. *Icarus* **117**, 287–312 (1995).
27. Karjalainen, R. & Salo, H. Gravitational accretion of particles in Saturn's rings. *Icarus* **172**, 328–348 (2004).
28. Daisaka, H., Tanaka, H. & Ida, S. Viscosity in a dense planetary ring with self-gravitating particles. *Icarus* **154**, 296–312 (2001).
29. Showalter, M. R. & Nicholson, P. D. Saturn's rings through a microscope—particle size constraints from the Voyager PPS scan. *Icarus* **87**, 285–306 (1990).
30. Porco, C. C. et al. Cassini Imaging Science: instrument characteristics and anticipated scientific investigations at Saturn. *Space Sci. Rev.* **115**, 363–497 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank E. Baker for help with data reduction, and P. Nicholson for discussions. We acknowledge support from JPL, the Cassini project, NASA's Planetary Geology and Geophysics program, and the UK Particle Physics and Astronomy Research Council.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.S.T. (matthewt@astro.cornell.edu).

Controlled multiple reversals of a ratchet effect

Clécio C. de Souza Silva^{1,†}, Joris Van de Vondel¹, Mathieu Morelle¹ & Victor V. Moshchalkov¹

A single particle confined in an asymmetric potential demonstrates an anticipated ratchet effect by drifting along the ‘easy’ ratchet direction when subjected to non-equilibrium fluctuations^{1–3}. This well-known effect can, however, be dramatically changed if the potential captures several interacting particles. Here we demonstrate that the inter-particle interactions in a chain of repelling particles captured by a ratchet potential can, in a controllable way, lead to multiple drift reversals, with the drift sign alternating from positive to negative as the number of particles per ratchet period changes from odd to even. To demonstrate experimentally the validity of this very general prediction, we performed transport measurements on a.c.-driven vortices trapped in a superconductor by an array of nanometre-scale asymmetric traps. We found that the direction of the vortex drift does undergo multiple reversals as the vortex density is increased, in excellent agreement with the model predictions. This drastic change in the drift behaviour between single- and multi-particle systems can shed some light on the different behaviour of ratchets and biomembranes⁴ in two drift regimes: diluted (single particles) and concentrated (interacting particles).

Contrary to what intuition could perhaps tell us, particles in a ratchet potential can, under special conditions, move preferentially along the direction where the potential barriers are steeper, that is, along the ‘hard’ direction. This effect can be crucial in the design of artificial ratchet-based devices capable of shuttling or separating—for instance, colloidal suspensions⁵ and DNA molecules⁶. In theory, an inversion in the drift direction of a single-particle brownian ratchet is predicted to occur for non-zero thermal noise when the excitation frequency exceeds a certain critical value, which is usually high and very sensitive to the model parameters⁷. In a system of many weakly interacting particles, this effect can, however, be strongly reduced when the particle density is increased⁸. Drift inversions have also been observed in mixtures of interacting brownian particles⁹ and in chaotic underdamped ratchets at zero thermal noise¹⁰. Here we show that, in a system of strongly interacting particles in a ratchet potential, the drift direction undergoes controllable multiple sign inversions as a function of particle density. These inversions do not require thermal or chaotic noise, or high excitation frequencies or a mixture of particles. Rather, they are ruled deterministically by the internal degrees of freedom of the system, providing a simple way to tune the drift direction of ratchet devices.

We consider a one-dimensional (1D) system of particles interacting via the pair potential $V_{\text{int}}(r) = -E_0 \ln(r)$, with r the pair separation and E_0 the relevant energy scale, in the double-well ratchet potential

$$U_p(x) = -U_{p1}e^{-\sin^2(\pi x)/2\sin^2(\pi R)} - U_{p2}e^{-\sin^2(\pi(x-d))/2\sin^2(\pi R)} \quad (1)$$

where U_{p1} and U_{p2} determine the depth of the stronger and weaker wells, respectively, which are separated by a distance $d = 0.36$ and have width $R = 0.15$, and x is the position. All lengths are in units of the ratchet period a . The dynamics of the chain is studied by

molecular dynamics simulations of the Langevin equations,

$$m\ddot{x}_i = -\eta\dot{x}_i - \sum_j \nabla V_{\text{int}}(x_i - x_j) - \nabla U_p(x) + F + \Gamma_i \quad (2)$$

where m is the mass of the particles, η the friction coefficient, F the external drive, and Γ_i the gaussian thermal noise¹¹. Hereafter we adopt $m = 1$ and $\eta = 16$, which corresponds to strongly overdamped dynamics.

Figure 1a shows density plots of the effective asymmetry in the critical forces for drifting the particles to the positive (F_{c+}) and to the negative (F_{c-}) direction, $\alpha_{\text{eff}} = 1 - F_{c+}/F_{c-}$. The sign of α_{eff} determines the preferential drift direction—positive (‘easy’) direction for $\alpha_{\text{eff}} > 0$ and negative (‘hard’) direction for $\alpha_{\text{eff}} < 0$ —whereas its magnitude is a measure of the ratchet efficiency. The plots are presented in the $\tilde{U}_{p1}$ – β plane ($\tilde{U}_{p1} = U_{p1}/E_0$ determines the potential strength relative to inter-particle interactions and $\beta = U_{p1}/U_{p2}$ determines the potential asymmetry) for occupation number $n = 1, 2, 3$ and 4 particles per ratchet period and for zero noise ($T = 0$). For $n = 1$ the particles are more easily driven to the usual positive direction ($\alpha_{\text{eff}} > 0$) and, except for $\tilde{U}_{p1} < 1$ (where the potential cannot trap the chain effectively), α_{eff} varies only with β . However, for $n > 1$, α_{eff} has a much richer dependence on the ratchet potential parameters, assuming either positive or negative values with comparable intensity. Particularly, there is a large region of the phase diagrams ($\beta > 0.56$ and moderate pinning strengths) where α_{eff} is always positive for odd n and negative for even n . In this region, particles distribute evenly between the weak and strong pinning sites for even n , whereas for odd n the strong traps capture one particle more than the weak ones (Fig. 1b).

A simple way to understand this interesting effect is to consider each local well in a ratchet period as being characterized by the effective energies E_1 (‘strong’ well) and E_2 (‘weak’ well). For $n = 0$, the strong trap yields a lower energy than the weak one ($E_1(0) < E_2(0)$). For $n = 1$, the particle occupies the strong well, raising its effective energy enough to surpass the energy of the (empty) trap 2 ($E_1(1) > E_2(0)$). A second particle will find a stable position at trap 2, then raising its energy above $E_1(1)$ ($E_1(1) < E_2(1)$). By increasing n even further, the rise in the effective energies proceeds following a brick-wall tiling pattern, with the particles populating each trap alternately. Thus, for n even, there is necessarily a smaller energy input required to move one particle from trap 2 (across the small inner energy barrier) to trap 1 (as $E_1(n_1) < E_2(n_2 = n_1)$), whereas for odd n a transition from 1 to 2 is favoured (as $E_1(n_1) > E_2(n_2 = n_1 - 1)$). Particles that are the closest to the inner energy barrier are the natural candidates to undergo such transitions. In this sense, these particles are the most weakly pinned ones.

As a first demonstration of the ratchet mechanism in this system, we excite the particles with an a.c. square-wave drive with an amplitude just above the threshold force (defined as $F_{\text{thresh}} = \min(F_{c+}, F_{c-})$) of the corresponding chain and a very low frequency (adiabatic drive). As illustrated in Fig. 1c, the motion of the weakly

¹INPAC—Institute for Nanoscale Physics and Chemistry, Nanoscale Superconductivity and Magnetism Group, Katholieke Universiteit Leuven, Celestijnenlaan 200 D, B-3001 Leuven, Belgium. [†]Present address: Departamento de Física, Universidade Federal de Pernambuco, 50670-901 Recife-PE, Brazil.

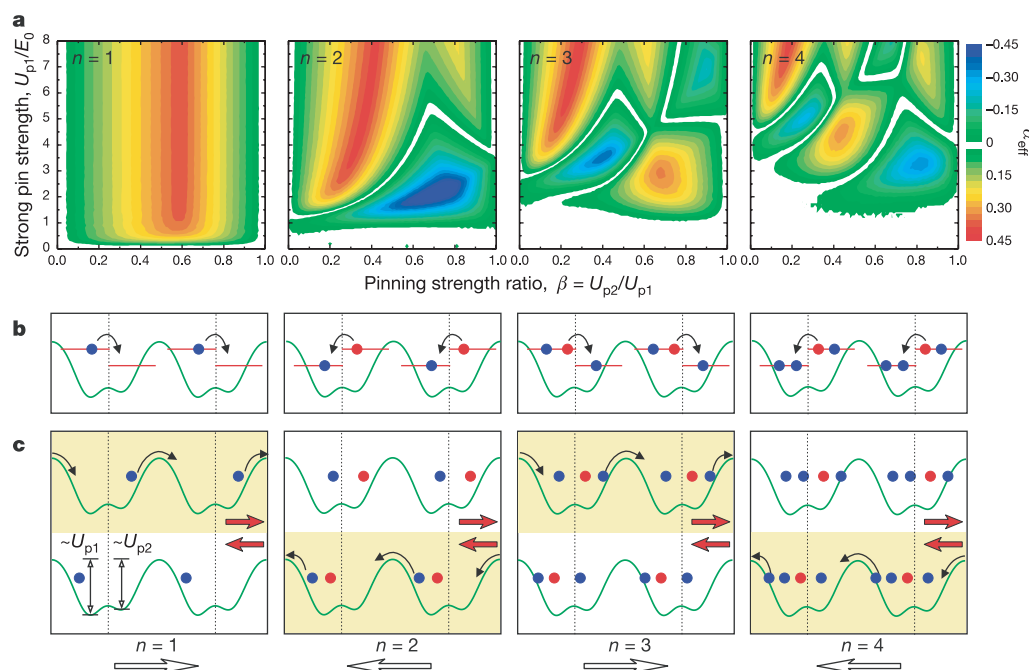


Figure 1 | Effective asymmetry and schematic demonstration of the ratchet mechanism. **a**, Density plots of the effective asymmetry $\alpha_{\text{eff}} = 1 - F_{c+}/F_{c-}$ as a function of the ratchet parameters $\tilde{U}_{p1} = U_{p1}/E_0$ and $\beta = U_{p2}/U_{p1}$ (<1) for $n = 1$ to 4 (see text for details). The potential has one minimum per period for $\beta < 0.56$ and two local minima per period for $\beta > 0.56$. We forced white shading for $\alpha_{\text{eff}} = 0$ to enhance the contrast between the positive and negative drift phases. The critical forces F_{c+} and F_{c-} were obtained by varying the driving force quasi-statically to the positive and negative directions respectively and assuming as a criterion for macroscopic drift that all particles travel a distance of at least one ratchet period. **b**, Diagram of the equilibrium configurations for $n = 1$ to 4 obtained by annealing the chain down to zero temperature with $U_{p1}/E_0 = 3.2$ and $U_{p2} = 0.9U_{p1}$, which generates a double-well ratchet potential (green curves). The relative characteristic energies of each pinning well (the energy

of a well plus the energy of the trapped vortices) and their respective occupancies are schematically represented. Owing to the excess in energy, one particle in a higher-energy trap is 'looser' than the others. Such particle (marked in red) is the most favourable for performing a transition (black arrows) through the inner energy barrier. **c**, Schematic demonstration of the ratchet mechanism when the chain is excited by an a.c. square-wave force with an amplitude just above the threshold force. Red arrows indicate the force direction. Yellow backgrounds highlight macroscopic motion of the chain in the corresponding drive direction, whereas white backgrounds indicate that the chain is at rest (pinned). The macroscopic drift is triggered by a transition of the most weakly pinned particle to the next available pinning site, as indicated in **b**. In sequence, one particle in this site is knocked out to the next ratchet period (as indicated by the black arrows), starting up motion of the whole chain.

pinned particle across the inner energy barrier triggers the whole ratchet mechanism (see also the Supplementary Videos). After transition, this particle 'overpopulates' the target well, which then releases another particle to the next ratchet period. When the drive inverts its sign, no motion is detected. This produces a net rectified motion with positive direction for odd n and negative direction for even n . To study in more detail the dependence of rectification on n , we calculated the net velocity of the chain as a function of n and $\tilde{U}_{p1}$

for a constant sinusoidal a.c. bias (Fig. 2). The result demonstrates remarkable sign reversals every time n approaches an integer value. We have also tested these predictions for the well-known double-sine potential⁷ (sketches of this and the double-well potentials are provided in Supplementary Fig. S1). In a large range of the potential parameters multiple reversals were also observed (compare Supplementary Fig. S2). To evaluate further the generality of our findings, we performed similar calculations for other friction values

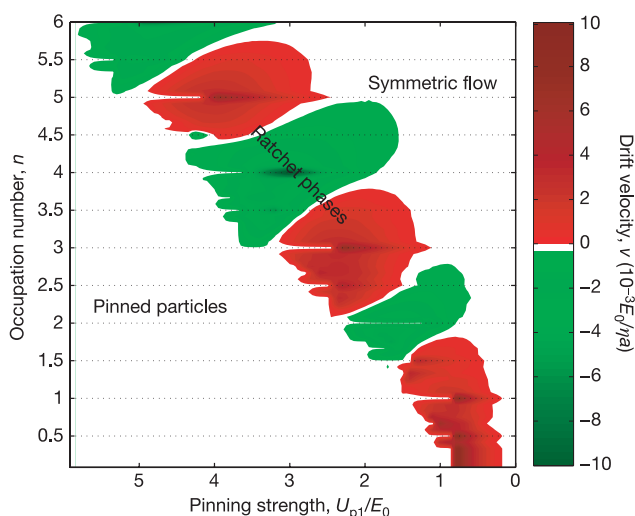


Figure 2 | Net drift velocity of the chain as a function of occupation number and pinning strength. The chain was adiabatically excited with a sinusoidal forcing of amplitude $F_{\text{ac}} = 3E_0/a$ and frequency $f = 5 \times 10^{-7} m/\eta$ at zero thermal noise (here a is the ratchet period, m is the particle mass, and η is the friction coefficient). The simulation cell comprises 12 periods of the double-well ratchet potential ($U_{p1}/E_0 = 3.2$ and $U_{p2} = 0.9U_{p1}$) with periodic boundary conditions. The U_{p1}/E_0 axis is presented in decreasing order for further comparison with Fig. 3c. The white areas may correspond to a pinned phase, where particles just oscillate inside the traps, or a symmetric moving phase. Also indicated are the ratchet phases exhibiting multiple sign inversions. The chain is rectified with maximum efficiency at integer and half-integer occupation numbers.

down to $\eta = 2$, which corresponds to the (regular) underdamped regime. In general, the results are very similar to those presented in Figs 1 and 2.

To test these predictions experimentally, we performed transport measurements of a.c.-driven vortices in a nanostructured superconducting film with an array of asymmetrical pinning sites. Vortices are whirlpools of current carrying one quantum of magnetic flux ($\Phi_0 = 2.07 \times 10^{-15}$ Wb) that repel each other and are attracted by microholes (termed antidots) in a superconductor¹². The vortex

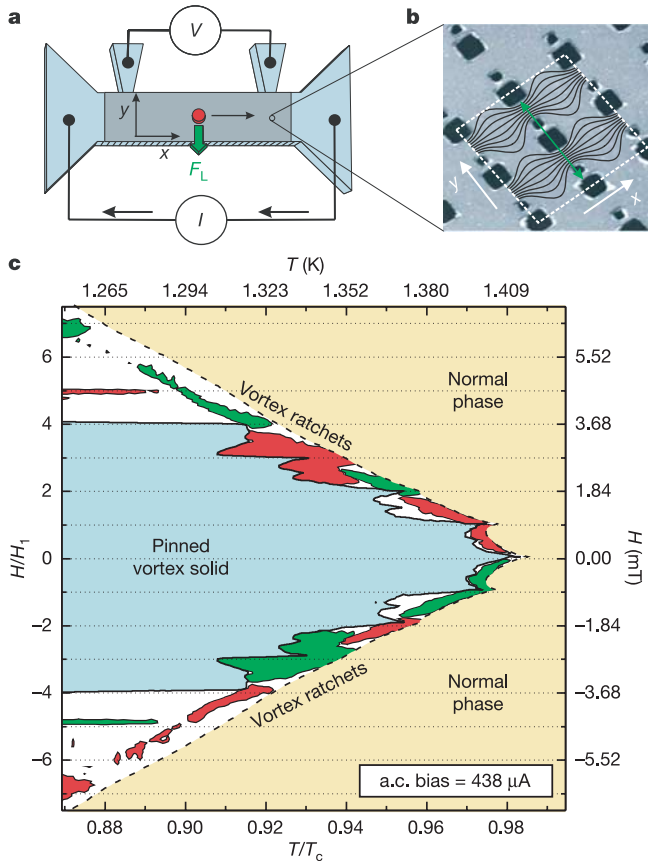


Figure 3 | Sample geometry and phase diagram of the vortex ratchet effect. **a**, The external magnetic field H generates a certain vortex distribution in the film. A vortex (shown schematically out of scale) is driven along the y direction by a Lorentz force $F_L = (J \times v) \Phi_0 d$ generated by an electrical current density J applied in the x direction (v is the normalized vortex circulation, parallel to H). If F_L is strong enough, vortices start moving along the drive with mean velocity v generating a voltage drop $V = L(v \times H) \cdot \hat{e}_z$ across a distance L . **b**, Atomic force micrograph of the double-antidot array (with period $a_p = 1.5 \mu\text{m}$). The big and small antidots are $600 \times 600 \text{ nm}^2$ and $300 \times 300 \text{ nm}^2$ in size and separated by a 90-nm-thick superconducting wall. Details of the sample preparation and characteristics are given elsewhere¹³. The streamlines of the applied electrical current (shown schematically) are substantially denser between antidots than in the interstitial positions, forcing the vortices to move preferentially along the antidot rows. As the driving Lorentz force is always perpendicular to these lines, motion occurs along the broken symmetry (y) direction. **c**, H - T dynamical phase diagram at an a.c. current $I(t) = I_{ac} \sin(2\pi ft)$, with $I_{ac} = 438 \mu\text{A}$ ($J_{ac} = 3.95 \times 10^3 \text{ A cm}^{-2}$) and $f = 1 \text{ kHz}$. H_1 is the first matching field and T_c is the superconducting critical temperature. Between the pinned vortex solid and normal phases (compare Fig. 4), the voltage is dominated by vortex motion. The green and red areas correspond to positive and negative V_{dc} respectively. In the white areas, vortex motion is symmetric ($V_{dc} \approx 0$) within the experiment accuracy. Note that the rectification mechanism is insensitive to the vortex polarity, since the interaction of vortices or antivortices with a microhole is the same. This leads to a symmetric net d.c. velocity, $v_{dc}(H) = v_{dc}(-H)$, which then results in an antisymmetric d.c. voltage, $V_{dc}(H) = -V_{dc}(-H)$.

density can be varied continuously by applying an external magnetic field H and, as shown in Fig. 3a, their dynamics can be probed by measuring the voltage-current characteristics of the sample. Our sample is an Al film (with critical temperature $T_c = 1.437 \text{ K}$) patterned by electron-beam lithography with a square array (with period $a_p = 1.5 \mu\text{m}$) of neighbouring big and small antidots placed close to each other, thus generating an asymmetric double-well vortex trap with broken symmetry along the y direction only (Fig. 3b). As we have recently demonstrated, such a configuration provides efficient rectification of vortex motion at low fields^{13,14}.

Our experiment is carried out as follows: an oscillating driving force (generated by a sinusoidal transverse electrical current) is applied along the direction of broken symmetry, and the vortex motion in this direction is probed by measuring the transverse voltage (Fig. 3a). A phase diagram of vortex motion was obtained by detailed measurements of the root-mean-square and d.c. voltages (V_{rms} and V_{dc} respectively) across the sample (Fig. 3c). In the pinned vortex solid (PVS) phase, the applied current is not high enough to drive vortices out of their equilibrium positions. At some vortex densities (rational multiples of the first matching field, $H_1 = \Phi_0/a_p^2 = 0.92 \text{ mT}$, where the number of vortices matches the number of double-traps), vortices assemble in a very stable lattice commensurate with the pinning array^{15,16}. These special configurations enhance the critical current, producing the sharp re-entrances of the PVS phase at integer and half-integer matching fields. The moving vortex phase is dominated by ratchet dynamics exhibiting multiple drift reversals. From the first up to the fifth matching fields, the direction of net vortex motion changes its sign alternately, resembling the sign inversions of

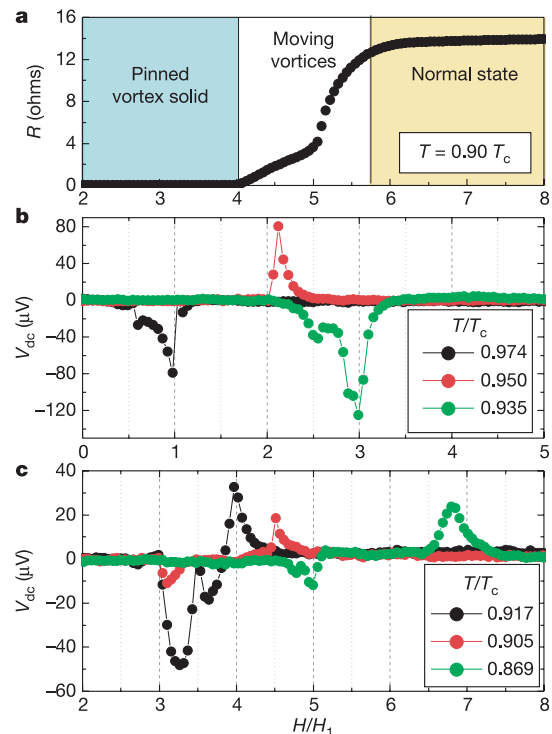


Figure 4 | Magnetoresistance and magnetic field dependence of the ratchet effect for an a.c. bias $I_{ac} = 438 \mu\text{A}$. **a**, By performing a.c. magnetoresistance ($R(H) = \sqrt{2} V_{rms}/I_{ac}$) measurements, we determine the boundaries between the pinned vortex solid, moving vortices and normal phases. When vortices start moving, R increases towards the normal state resistance, R_n . The moving vortex phase is then bounded by the criteria $R = 10^{-5} R_n$ for the onset of vortex motion, and $R = 0.90 R_n$ for the destruction of superconductivity. In **b** and **c**, the measured d.c. voltage, V_{dc} , is plotted against magnetic field for several temperature values. The curves exhibit multiple sign reversals of the d.c. voltage with maxima and minima close to integer and half-integer matching fields.

the chain drift in our 1D model (Fig. 2). Thermal fluctuations are negligible in our sample, because the pinning energy is typically much higher than kT ($U_p \approx 10^2 - 10^3 kT$, for $T/T_c = 0.98 - 0.88$). Hence, the vortex dynamics is essentially deterministic. The temperature does however play an important role in determining the pinning efficiency of an antidot. At temperatures very close to T_c , vortices are bigger than the antidots, which then become less effective pinning centres. At lower temperatures, vortices become smaller and interact more strongly with the antidots¹². In this sense, decreasing the temperature plays the role of increasing the pinning strength.

Sign reversal in a vortex ratchet has been reported previously for an array of triangular magnetic dots¹⁷. One single reversal was observed to take place gradually as the number of vortices increased above the corresponding saturation of the dots (three vortices per dot). This was interpreted as the effect of interstitial vortices moving in an inverted ratchet potential produced by the interactions with the trapped vortices. The multiple sign reversals observed in our experiment cannot be explained by the inverted ratchet effect of interstitial vortices. Rather, owing to the strong enhancement of the current density between the antidots (Fig. 3b), vortices tend to move in 1D channels along the antidot rows. These channels should however saturate at a high enough vortex concentration, the excess vortices being forced to move along the interstitial positions. It is also noteworthy that vortices are collective excitations; their cores can be deformed and merged into one another at extreme conditions. Consequently, one must be cautious when modelling vortices as hard particles. Nonetheless, the agreement of the experimental results with the model predictions is quite good, which suggests that the model is able to capture the main physics of the observed vortex ratchet effects. These multiple sign reversals provide a new tool for controlling and manipulating the motion of magnetic flux quanta in superconductors. Finally, we stress that our findings have a very general character and are also relevant to other ratchet systems of interacting particles, like charged colloidal suspensions in ratchet-like microtubules and ions in the selectivity filter of ion channels in cell membranes⁴.

Received 17 June 2005; accepted 18 January 2006.

1. von Smoluchowski, M. Experimentell nachweisbare, der üblichen Thermodynamik widersprechende Molekularphänomene. *Phys. Z.* **13**, 1069–1080 (1912).

2. Feynman, R. P., Leighton, R. B. & Sands, M. *The Feynman Lectures On Physics* Ch. 46 Vol. 1 (Addison-Wesley, Reading, Massachusetts, 1966).
3. Magnasco, M. O. Forced thermal ratchets. *Phys. Rev. Lett.* **71**, 1477–1481 (1993).
4. Morais-Cabral, J. H., Zhou, Y. & MacKinnon, H. Energetic optimization of ion conduction rate by the K^+ selectivity filter. *Nature* **414**, 37–42 (2001).
5. Matthias, S. & Müller, F. Asymmetric pores in a silicon membrane acting as massively parallel brownian ratchets. *Nature* **424**, 53–57 (2003).
6. Bader, J. S. *et al.* DNA transport by a micromachined Brownian ratchet device. *Proc. Natl Acad. Sci. USA* **96**, 13165–13169 (1999).
7. Bartussek, R., Hänggi, P. & Kissner, J. G. Periodically rocked thermal ratchets. *Europhys. Lett.* **28**, 459–464 (1994).
8. Derényi, I. & Vekse, T. Cooperative transport of Brownian particles. *Phys. Rev. Lett.* **75**, 374–377 (1995).
9. Savel'ev, S., Marchesoni, F. & Nori, F. Controlling transport in mixtures of interacting particles using Brownian motors. *Phys. Rev. Lett.* **91**, 010601 (2003).
10. Mateos, J. L. Chaotic transport and current reversal in deterministic ratchets. *Phys. Rev. Lett.* **84**, 258–261 (2000).
11. Riskin, H. *The Fokker-Planck Equation* Ch. 11 (Springer, New York, 1984).
12. Blatter, G., Feigel'man, M. V., Geshkenbein, V. B., Larkin, A. I. & Vinokur, V. M. Vortices in high-temperature superconductors. *Rev. Mod. Phys.* **66**, 1125–1388 (1994).
13. de Souza Silva, C. C., Van de Vondel, J., Zhu, B. Y., Morelle, M. & Moshchalkov, V. V. Vortex ratchet effects in films with a periodic array of antidots. *Phys. Rev. B* **73**, 014507 (2006).
14. Van de Vondel, J., de Souza Silva, C. C., Zhu, B. Y., Morelle, M. & Moshchalkov, V. V. Vortex-rectification effects in films with periodic asymmetric pinning. *Phys. Rev. Lett.* **94**, 057003 (2005).
15. Fiory, A. T., Hebard, A. F. & Somekh, S. Critical currents associated with the interaction of commensurate flux-line sublattices in a perforated Al film. *Appl. Phys. Lett.* **32**, 73–75 (1977).
16. Baert, M., Metlushko, V. V., Jonckheere, R., Moshchalkov, V. V. & Bruynseraede, Y. Composite flux-line lattices stabilized in superconducting films by a regular array of artificial defects. *Phys. Rev. Lett.* **74**, 3269–3272 (1995).
17. Villegas, J. E. *et al.* A superconducting reversible rectifier that controls the motion of magnetic flux quanta. *Science* **302**, 1188–1191 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Silhanek for the critical reading of our paper and S. Raedts for taking the AFM micrographs. This work was supported by the K.U. Leuven Research Fund GOA and FWO programmes. C.C.d.S.S. was supported by CNPq, an Agency of the Brazilian Government.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to V.V.M. (Victor.Moshchalkov@fys.kuleuven.be).

***Ab initio* determination of solid-state nanostructure**

P. Juhás¹, D. M. Cherba², P. M. Duxbury¹, W. F. Punch² & S. J. L. Billinge¹

Advances in materials science and molecular biology followed rapidly from the ability to characterize atomic structure using single crystals^{1–4}. Structure determination is more difficult if single crystals are not available⁵. Many complex inorganic materials that are of interest in nanotechnology have no periodic long-range order and so their structures cannot be solved using crystallographic methods⁶. Here we demonstrate that *ab initio* structure solution of these nanostructured materials is feasible using diffraction data in combination with distance geometry methods. Precise, sub-ångström resolution distance data are experimentally available from the atomic pair distribution function (PDF)^{6,7}. Current PDF analysis consists of structure refinement from reasonable initial structure guesses^{6,7} and it is not clear, a priori, that sufficient information exists in the PDF to obtain a unique structural solution. Here we present and validate two algorithms for structure reconstruction from precise unassigned interatomic distances for a range of clusters. We then apply the algorithms to find a unique, *ab initio*, structural solution for C₆₀ from PDF data alone. This opens the door to sub-ångström resolution structure solution of nanomaterials, even when crystallographic methods fail.

Powerful direct imaging methods, such as scanning tunnelling microscopy, transmission electron microscopy and, more recently, lensless imaging⁸, are available to characterize the structure of nanomaterials; however, they do not yield the high precision three-dimensional structural information traditionally obtained using crystallographic methods. The effort towards high accuracy structure determination is driven by the fact that even small changes in interatomic bond lengths can have a marked effect on the properties of solid state materials. For example, the key polaron distortion in giant magnetoresistive materials is of the order of one-tenth of an ångström⁹. Extended X-ray absorption fine structure analysis yields high precision values for the local environment of atoms in nanoparticles¹⁰ but not a complete structure. Nuclear magnetic resonance (NMR) in combination with distance geometry methods is critical to structure solution of proteins¹¹, particularly in the absence of protein single crystals. However, nuclear Overhauser effect distances used in protein NMR analysis have low resolution, with uncertainties of the order of one ångström¹². The distance lists extracted from PDF data of nanostructured solids have high resolution, with uncertainties of the order of a few hundredths of an ångström in the atomic separations. However, despite PDFs of materials being measured for almost 75 years (ref. 7), *ab initio* structure solution from such data has not been previously demonstrated. Here we present and validate several algorithms for structure solution from such high precision, but unassigned, distance lists.

The PDF method was traditionally applied to the study of glasses and liquids¹³ but more recently has also successfully yielded information about atomic-scale structures of nanosized materials^{6,10,14,15}. For example, the structure of ZnS nanoparticles was found to be significantly modified from the expected sphalerite structure that had been inferred from transmission electron microscopy observations¹⁴.

Another important area of PDF application is nanostructured materials that have nanoscale inhomogeneities within a bulk matrix⁶. Atomic arrangements in these materials are well ordered locally, but are not long-range ordered and cannot be solved using crystallographic methods. PDF data are readily obtained using neutron and X-ray powder diffraction measurements, where area X-ray detectors allow remarkably rapid data acquisition¹⁶. Previously, analysis of PDF data has relied on known starting models¹⁴ or good structural analogues, and has used a trial-and-error approach^{6,17}, which is often a laborious process. Alternative methods such as reverse Monte Carlo¹⁸, empirical potential structure refinement¹⁹ and experimentally constrained molecular relaxation²⁰ are successful on highly disordered materials and provide a pool of candidate structures consistent with the data, but have not been used to reconstruct the structures of well ordered nanomaterials.

The PDF data from a single element system contains a simple unsorted list of the atomic distances present in the cluster without any orientational or three-body information. Reconstruction of structure from noisy or incomplete distances is computationally hard^{21,22} even when assignment of lengths to atom pairs is available, as is usually the case in protein structure solution using NMR. The distances extracted from PDF data are much more precise; however, the lengths are unassigned as the pair of atoms contributing to each distance is not known. Nevertheless, we find that a unique and efficient structure solution is possible from unassigned ideal distances for a wide range of clusters, including platonic solids, finite lattices of different symmetry, the C₆₀ 'buckyball' and Lennard-Jones minimum-energy clusters^{23,24}. More remarkably, we found that *ab initio* structure determination is also possible using distances extracted from experimental neutron PDF data for fullerenes.

The *n*-atom Lennard-Jones (LJ-*n*) cluster is the ground-state configuration of *n* atoms assuming a Lennard-Jones pair potential acting between all the atoms, and is a standard benchmark system for new optimization methods^{23–25}. We have used the interatomic distances occurring in these structures as the target distances for testing various distance geometry algorithms. The cost function that we optimize is the variance between the model distances and the target distances, namely $\text{var}(d) = \frac{1}{N_p} \sum_{k=1}^{N_p} (d_k^m - d_{l(k)}^e)^2$, where $N_p = N(N-1)/2$ is the number of atom pairs in the cluster, d_k is the interatomic distance of atom pair *k*, while the suffix *m* indicates the model and the suffix *e* indicates the experimental or target value. When $\text{var}(d) = 0$, the fit is exact. The most difficult computational aspect of this problem is correctly assigning the distances between model atom pairs *k* to target distances *l*(*k*). We first tried a simulated annealing approach²⁶, which was successful in finding the correct small clusters from unassigned distance data. However, this method failed for anything more complicated than a 20-atom cluster. This is presumably due to the rugged topology of the potential ($\text{var}(d)$) surface.

Genetic or evolutionary algorithms have been very successful in finding the ground state of many types of clusters using theoretical interatomic potentials^{23,25,27}. Based on these papers, we have developed

¹Department of Physics and Astronomy, ²Department of Computer Science and Engineering, Michigan State University, East Lansing, Michigan 48824, USA.

Table 1 | Performance comparison of distance geometry algorithms

Shape	Success rate	CPU time (s)			
		Average	Standard deviation	Minimum	Maximum
Genetic algorithm					
C ₆₀	35/62 (56%)	6,330	8,200	1,100	35,150
LJ-20	100/100 (100%)	39	75	7.3	620
LJ-38	103/113 (91%)	880	2,500	76	14,300
LJ-60	23/58 (40%)	10,080	12,500	1,200	48,100
LJ-88	35/43 (81%)	36,800	25,800	14,600	122,000
Liga algorithm					
C ₆₀	100/100 (100%)	1.1	0.74	0.17	3.7
LJ-20	100/100 (100%)	16	9.3	2.7	56
LJ-38	100/100 (100%)	3.0	2.8	0.1	14
LJ-60	100/100 (100%)	690	290	180	1,460
LJ-88	100/100 (100%)	2,650	1,800	937	15,200

All shapes were solved using unassigned ideal distances. The convergence criterion was $\text{var}(d) < 10^{-4} \text{ \AA}^2$. CPU (central processing unit) times were measured on an Intel Pentium 4, 2.66 GHz Linux PC. All runs were terminated and declared unsuccessful if not converged after 36 h. LJ-*n* indicates a Lennard-Jones cluster of *n* atoms.

a genetic algorithm for solving the unassigned distance geometry problem (see Methods section). This algorithm usually finds structures with relatively small $\text{var}(d)$ even for large structures. It also successfully found the correct C₆₀ and LJ-*n* clusters up to 150 atoms from ideal distance tables; however, it was relatively slow and unreliable for larger Lennard-Jones structures (see Table 1). To improve efficiency and accuracy, we developed a novel algorithm which grows large clusters by adding atoms to a population of high quality subclusters. This algorithm incorporates a strategy for backtracking and updating populations of high quality clusters at each size (Fig. 1, and Supplementary Video 1), which is inspired by promotion and relegation in sport—such as occurs in European soccer leagues like La Liga in Spain (see Methods). Hereafter we refer to this procedure as the Liga algorithm. Both the genetic algorithm and the Liga algorithm were tested on ideal distance data from simple geometrical shapes, LJ-*n* clusters and the ideal buckyball, and some of the timing results are presented in Table 1. In all of the cases we have tried, the Liga algorithm performed better, both in the quality of the solution and the speed of convergence; this was the case for both highly symmetric structures such as fullerenes, and for lower-symmetry structures such as triclinic finite lattices or LJ-*n* clusters. The Liga algorithm has been developed for nanostructure determination by taking advantage of the nature of the data in the PDF. However, we are exploring the possibility that this particular combination of strategies, which involve the subunit buildup aspect of dynamic programming and tournaments used in genetic algorithms, has broader application in the field of hard computational problems.

To be of interest to real materials, it is essential to extract and use distances from measured PDF data. We demonstrate that this is possible using room-temperature neutron PDF data measured on solid C₆₀ as shown in Fig. 2a. The raw data contain both the probabilities of intramolecular distances (sharp peaks at interatomic separations, *r*, below 7.1 Å) and the particle–particle correlations. The particle–particle correlation function was estimated using the approach of ref. 28, and subtracted from the data. The intraparticle correlation function, which is the focus of this work, was then converted to the radial distribution function shown in Fig. 2b. Distances were extracted from the data by identifying the positions of peak maxima or of shoulders to peaks (see Methods). The resulting distance table is distorted from the ideal distance table because of noise and also because of uncertainties due to peak overlap in the PDF data. For example, owing to noise and peak overlap, the data derived table has 18 instead of 21 unique distances and 184 second neighbour distances compared to 180 for the ideal table. The Liga solution to this table leads to a defective structure, as shown in Fig. 2c, that has a lower $\text{var}(d)$ than the ideal buckyball, indicating that errors of multiplicity prevent convergence to the correct structure. To

account for these errors, instead of fitting the ‘tight’ table, where the number of distances is exactly equal to the number of pairs in the 60-atom cluster, we fit a ‘loose’ table that allowed a greater multiplicity at each distance. We found that the Liga algorithm converged to the correct structure (Fig. 2d) when we included at least a 10% looseness in the multiplicity. A typical run time was about 1,200 s on an Intel Pentium 4, 2.66 GHz, Linux PC. In fact, surprisingly, we find that the C₆₀ molecule can be rapidly reconstructed even with a completely ‘loose’ table where the multiplicity of each distance is allowed to be arbitrary.

Here we have demonstrated, to our knowledge for the first time, that sufficient information exists in experimental PDF data alone to reconstruct a rigid cluster such as C₆₀, and we present an efficient algorithm for making the reconstruction. In our initial implementation, no a priori knowledge about the system, such as symmetry, chemical or bonding information, was needed to find the solution. Extensions to the algorithm will be important in solving problems that are ill-conditioned in the sense that there is not enough information in the PDF data alone to result in a unique solution. The Liga algorithm can be straightforwardly extended to include chemical and physical constraints. For example, the cluster buildup procedure can utilize known bond lengths and bond angles. Another way of applying chemical knowledge during cluster buildup is to exclude unfeasible near neighbours, such as Na–Na or Cl–Cl pairs in sodium chloride, and to use known structure subunits (for example,

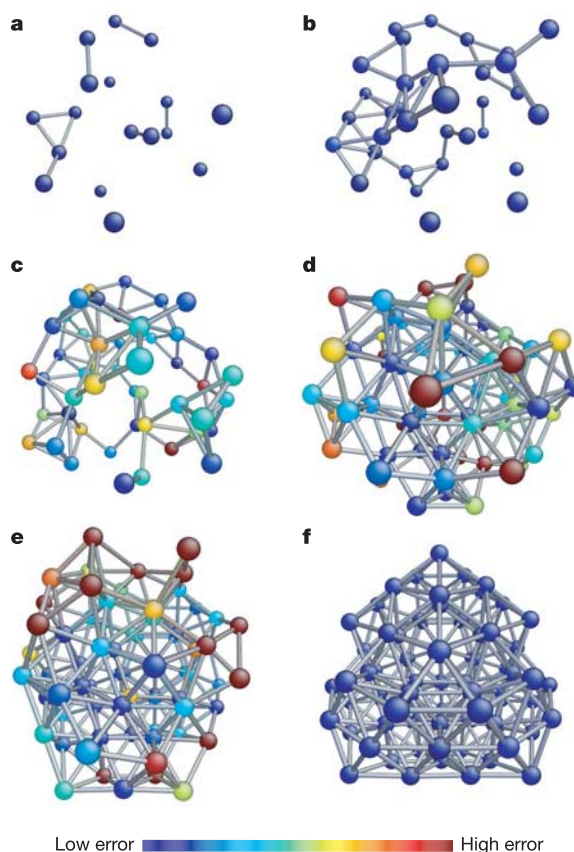


Figure 1 | Reconstruction of an LJ-88 cluster from unlabelled distances using the Liga algorithm. Atom colours denote contributions to the total error. **a, b,** The procedure starts by building partial clusters using only allowed distances. **c, d,** As more atoms are added the system becomes more constrained, leading to the appearance of high-error (red) atoms. **e, f,** Badly placed atoms are removed, allowing regrowth to the correct structure. The displayed clusters contain 17, 34, 51, 72, 79 and 88 atoms. The final cluster (**f**) is topologically identical to the target LJ-88 cluster with $\text{var}(d) = 2 \times 10^{-5} \text{ \AA}^2$.

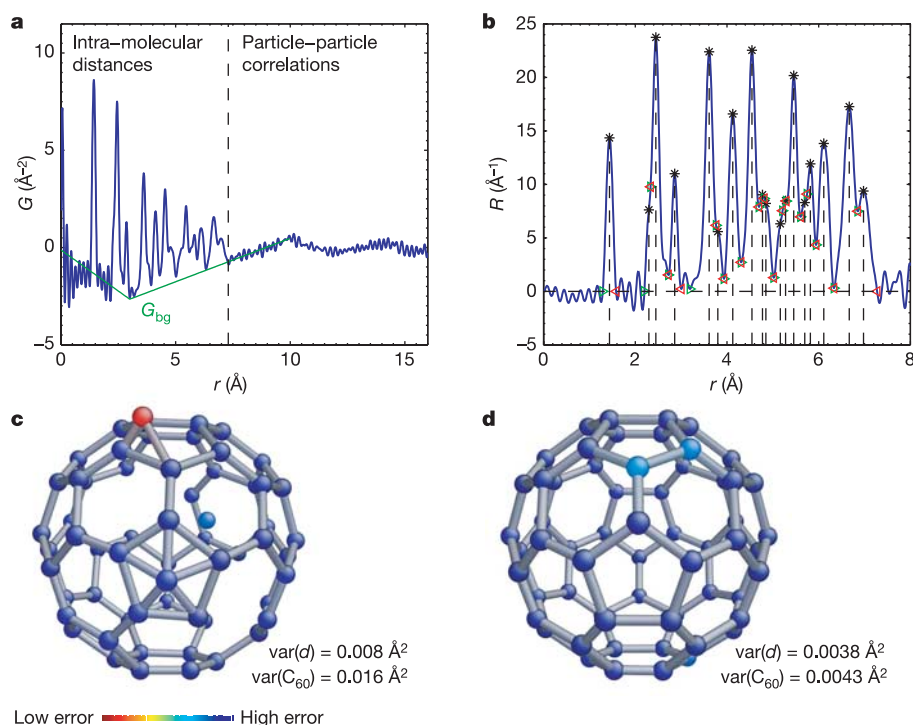


Figure 2 | Structure solution of fullerene from neutron PDF data.

a, Experimental pair distribution function, G , from solid C_{60} as a function of distance, r . The green line, G_{bg} , shows background arising from interparticle correlations. **b**, The background-subtracted data in the form of the radial distribution function, R . The experimental interatomic distances were obtained from the positions of peak maxima and shoulders (asterisks), and

their multiplicities were set in proportion to the peak areas, where green and red triangles denote integration limits. **c**, The C_{60} cluster derived from this 'tight' list of distances, and **d**, the cluster obtained from a 10% loose list. Here $\text{var}(C_{60})$ is the error of the ideal 'buckyball' with respect to the experimental data.

aromatic rings) as building blocks instead of single atoms. These extensions will allow larger and lower-symmetry clusters to be solved from imperfect data. Another particular advantage of the distance geometry approach described here is the ease with which data from several complementary experimental probes can be combined to constrain solutions. For example, extended X-ray absorption fine structure analysis and NMR provide measures of local distances that are chemically specific, though limited in range—highly complementary to the information in the PDF data. The extension to multi-element systems is straightforward. In the current version the cluster

buildup is gradually consuming all distances available in the target distance list. In the multi-element case the partial clusters will instead use up fractional amplitudes of the observed PDF peak intensities, since all distance counts are scaled by the scattering power of corresponding atom pairs.

Larger and lower-symmetry clusters will present special problems because of the dual factors that the information in an experimental PDF decreases owing to peak overlap and that the combinatorics of the problem increases. We currently do not know the fundamental limits on these aspects, but have successfully reconstructed from ideal data LJ-150 (193 unique distances) and a 112-atom supercell of distorted CeTe_3 (625 unique distances), among other problems currently under investigation in the group. It is promising that clusters of the order of 100 atoms and of moderate symmetry can be solved from PDF data alone. Clearly, structure solution of larger, lower-symmetry, clusters will rely on incorporating information from complementary data, and chemical and physical constraints. As many nanostructures can be described by subunits that are of the order of 100 atoms^{6,14}, the *ab initio* determination of their structure is now feasible.

METHODS

Liga algorithm. The Liga algorithm is illustrated in Fig. 3, which tracks reconstruction of an octahedron from its ideal list of 15 distances. The algorithm starts with a single atom, and the second atom is added at a randomly selected distance from the target list. The third position is found by constructing a triangle using two target distances, and additional atoms are added by constructing 4-vertex pyramids, while attempting to use only the allowed distances from the target list (Fig. 3a–c). However, there are many small clusters that use allowed target lengths, but are inconsistent with the target structure—for example, the tetrahedron shown in Fig. 3c is not part of an octahedron. Growth from these incorrect clusters eventually leads to an increase in the cost function (Fig. 3d) and the algorithm then has to backtrack to repair the faulty part of the cluster (Fig. 3e, f).

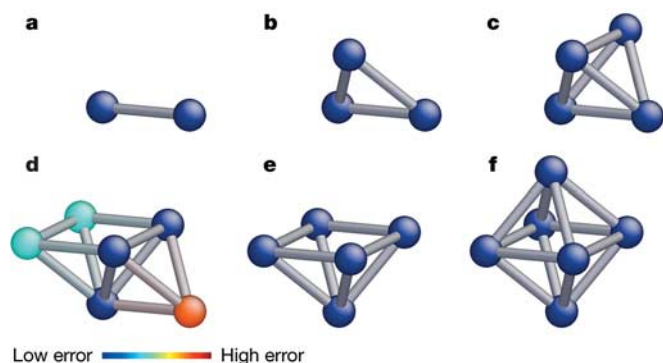


Figure 3 | Illustration of the Liga algorithm for an octahedron. **a**, The initial cluster consists of two atoms separated by a distance from the target list. **b**, **c**, More atoms are added while ensuring that the distance table is not violated. **d**, Because the tetrahedron is not part of an octahedron, addition of more atoms induces error with individual contributions indicated by the atom colours. **e**, The worst atom is removed, allowing the cluster to proceed to the correct solution (**f**).

Backtracking is carried out by first evaluating the individual atom contributions to the total error and removing the 'worst' atoms according to a stochastic procedure where the probability an atom is removed is proportional to its associated error contribution. The backtracking procedure itself is inspired by the concept of promotion and relegation in sports leagues, such as the European soccer leagues. An N -atom cluster has a set of N subclusters of size $n = 1, \dots, N$, and we keep a population of 10 clusters at each of these cluster sizes. In analogy with soccer leagues there are then N divisions and 10 teams in each division, where each team corresponds to a different cluster of n atoms. The competition of 'teams' is simulated by a random choice of winner and loser clusters, where the probability of winning is proportional to the reciprocal of the cost of the cluster. The 'champion' cluster tries to add as many atoms as possible, and so, unlike in soccer leagues, it may be promoted by several levels. The promoted cluster switches places with the most poorly performing cluster in the high division, which gets relegated to the lower division. On relegation, the cluster 'fires' its most poorly performing atom(s), thus decreasing its error and then has a chance of winning the lower division, acquiring new atom(s) and competing again in the higher level, hopefully with an improved 'game'. Relegation thus allows the cluster to recover from a dead-end search and correct the badly placed atoms, as illustrated in Fig. 3.

Genetic algorithm. We developed a genetic algorithm building on the work of Deaven and Ho²⁹ and Hartke²⁷. A key component of the algorithm is mating or crossover, where existing clusters are cut into equal-sized halves and the halves of different clusters are mated. Our genetic algorithm differs by its use of unlabelled distances for both the objective function and local search. In addition, a mutation operator based on complete relocation of an atom has been introduced. In the case of the buckyball reconstruction, we randomly initialized a population of 75 molecules and built successive replacement generations with the following probabilities for individual operators: reproduction 0.15, crossover 0.6, and combined mutation 0.25. Each molecule is then subjected to local search using 10 iterations. This local search utilizes a correction vector applied to each atom location derived by comparing the target distance table to specific atom pair distances. For the exact C_{60} distance data the algorithm finds a valid configuration in 260 generations.

Neutron PDF determination. The PDF method is described in detail elsewhere⁷. Powder diffraction data are collected over a wide range of momentum transfer, Q , using high-energy X-rays or neutrons. The data are corrected for experimental artefacts such as parasitic scattering, absorption and multiple scattering to obtain the structure dependent total scattering structure function, $S(Q)$, which is Fourier transformed according to $G(r) = \frac{2}{\pi} \int_{Q_{\min}}^{Q_{\max}} [S(Q) - 1] \sin(Qr) dQ$. The resulting pair distribution function, $G(r) = \frac{1}{N r(b)^2} \sum_{i \neq j} b_i b_j \delta(r - r_{ij}) - 4\pi r \rho_0$, is a scattering-length weighted measure of the probability of finding pairs of atoms in the material separated by the distance r , where N is the total number of atoms, b_i the scattering length of atom i , δ the Dirac function and ρ_0 the average number density. An example of the experimentally determined PDF from a C_{60} buckyball sample is shown in Fig. 2a. Data were obtained from room-temperature neutron scattering experiment measured at the Intense Pulse Neutron Source at Argonne National Laboratory²⁸. To extract the list of interatomic distances, the particle-particle background, $G_{bg}(r)$, was removed from $G(r)$ before converting to the radial distribution function, $R(r) = r[G(r) - G_{bg}(r)] = \frac{1}{N} \sum_{i \neq j} \frac{b_i b_j}{(b)^2} \delta(r - r_{ij})$, as shown in Fig. 2b. The distance multiplicities were obtained by integrating the $R(r)$ peaks, and scaling the total number of distances to the number of pairs in the 60-atom cluster.

Received 29 July; accepted 16 December 2005.

1. Friedrich, W., Knipping, P. & Laue, M. Interference appearances in x-rays. *Ann. Phys. (Berlin)* **41**, 971–988 (1913); reprinted from *Sitzb. K. Bayer. Akad. Wiss.* 303–322 (1912).
2. Bragg, W. H. & Bragg, W. L. The reflection of x-rays by crystals. *Proc. R. Soc. Lond. A* **88**, 428–438 (1913).
3. Perutz, M. F. *et al.* Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5 Å resolution, obtained by x-ray analysis. *Nature* **185**, 416–422 (1960).

4. Kendrew, J. B. *et al.* Structure of myoglobin: a three-dimensional Fourier synthesis at 2 Å resolution. *Nature* **185**, 422–427 (1960).
5. David, W., Shankland, K., McCusker, L. & Baerlocher, C. (eds) *Structure Determination from Powder Diffraction Data* (Oxford Univ. Press, Oxford, UK, 2002).
6. Billinge, S. J. L. & Kanatzidis, M. G. Beyond crystallography: the study of disorder, nanocrystallinity and crystallographically challenged materials. *Chem. Commun.*, 749–760 (2004).
7. Egami, T. & Billinge, S. J. L. *Underneath the Bragg Peaks: Structural Analysis of Complex Materials* (Pergamon, Oxford, UK, 2003).
8. Zuo, J. M., Vartanyants, I., Gao, M., Zhang, R. & Nagahara, L. A. Atomic resolution imaging of a carbon nanotube from diffraction intensities. *Science* **300**, 1419–1421 (2003).
9. Millis, A. J. Lattice effects in magnetoresistive manganese perovskites. *Nature* **392**, 147–150 (1998).
10. Zhang, H. Z., Gilbert, B., Huang, F. & Banfield, J. F. Water-driven structure transformation in nanoparticles at room temperature. *Nature* **424**, 1025–1029 (2003).
11. Clore, G. M. & Gronenborn, A. M. Determining the structures of large proteins and protein complexes by NMR. *Trends Biotechnol.* **16**, 22–34 (1998).
12. Nilges, M. & O'Donoghue, S. I. Ambiguous NOEs and automated NOE assignment. *Prog. Nucl. Magn. Reson. Spectrosc.* **32**, 107–139 (1998).
13. Wright, A. Diffraction studies of glass structure: the first 70 years. *Glass Phys. Chem.* **24**, 148–179 (1998).
14. Gilbert, B., Huang, F., Zhang, H., Waychunas, G. & Banfield, J. Nanoparticles: Strained and stiff. *Science* **305**, 651–654 (2004).
15. Page, K. *et al.* Direct observation of the structure of gold nanoparticles by total scattering powder neutron diffraction. *Chem. Phys. Lett.* **393**, 385–388 (2004).
16. Chupas, P. J. *et al.* Rapid acquisition pair distribution function analysis (RA-PDF). *J. Appl. Crystallogr.* **36**, 1342–1347 (2003).
17. Petkov, V. *et al.* Structure of nanocrystalline materials using atomic pair distribution function analysis: study of LiMoS_2 . *Phys. Rev. B* **65**, 092105 (2002).
18. McGreevy, R. L. & Pusztai, L. Reverse Monte Carlo simulation: a new technique for the determination of disordered structures. *Mol. Simul.* **1**, 359–367 (1988).
19. Soper, A. K. Empirical potential Monte Carlo simulation of fluid structure. *Chem. Phys.* **202**, 295–306 (1996).
20. Biswas, P., Tafel, D. & Drabold, D. A. Experimentally constrained molecular relaxation: The case of glassy GeSe_2 . *Phys. Rev. B* **71**, 054204 (2005).
21. Crippen, G. M. & Havel, T. F. *Distance Geometry and Molecular Conformation* (Wiley & Sons, New York, 1988).
22. Hendrickson, B. The molecule problem—exploiting structure in global optimization. *SIAM J. Optimiz.* **5**, 835–857 (1995).
23. Deaven, D., Tit, N., Morris, J. & Ho, K. Structural optimization of Lennard-Jones clusters by a genetic algorithm. *Chem. Phys. Lett.* **256**, 195–200 (1996).
24. Wales, D. J. & Scheraga, H. A. Review: Chemistry - global optimization of clusters, crystals, and biomolecules. *Science* **285**, 1368–1372 (1999).
25. Cai, W. S. & Shao, X. G. A fast annealing evolutionary algorithm for global optimization. *J. Comput. Chem.* **23**, 427–435 (2002).
26. Kirkpatrick, S., Gelatt, C. D. & Vecchi, M. P. Optimization by simulated annealing. *Science* **220**, 671–680 (1983).
27. Hartke, B. Global cluster geometry optimization by a phenotype algorithm with niches: Location of elusive minima, and low-order scaling with cluster size. *J. Comput. Chem.* **20**, 1752–1759 (1999).
28. Thorpe, M. F., Levashov, V. A., Lei, M. & Billinge, S. J. L. in *From Semiconductors to Proteins: Beyond the Average Structure* (eds Billinge, S. J. L. & Thorpe, M. F.) 105–128 (Kluwer/Plenum, New York, 2002).
29. Deaven, D. M. & Ho, K. M. Molecular-geometry optimization with a genetic algorithm. *Phys. Rev. Lett.* **75**, 288–291 (1995).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. P. Ramirez and R. C. Haddon for supplying the C_{60} sample. P.J. appreciates discussions with J. Bloch and E. S. Božin. P.M.D. acknowledges support from the Department of Energy (DOE) and S.J.L.B. from the NSF NIRT programme. Neutron data were collected at the GLAD instrument IPNS, which is funded by DOE.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.J.L.B. (billinge@pa.msu.edu).

Melting in the Earth's deep upper mantle caused by carbon dioxide

Rajdeep Dasgupta¹ & Marc M. Hirschmann¹

The onset of partial melting beneath mid-ocean ridges governs the cycling of highly incompatible elements from the mantle to the crust¹, the flux of key volatiles (such as CO₂, He and Ar)^{1,2} and the rheological properties of the upper mantle³. Geophysical observations^{4–6} indicate that melting beneath ridges begins at depths approaching 300 km, but the cause of this melting has remained unclear. Here we determine the solidus of carbonated peridotite from 3 to 10 GPa and demonstrate that melting beneath ridges may occur at depths up to 330 km, producing 0.03–0.3% carbonate liquid. We argue that these melts promote recrystallization and realignment of the mineral matrix, which may explain the geophysical observations. Extraction of incipient carbonate melts from deep within the oceanic mantle produces an abundant source of metasomatic fluids and a vast mantle residue depleted in highly incompatible elements and fractionated in key parent-daughter elements. We infer that carbon, helium, argon and highly incompatible heat-producing elements (such as uranium, thorium and potassium) are efficiently scavenged from depths of ~200–330 km in the upper mantle.

The mass of carbon stored in the mantle exceeds that in all other reservoirs of the global carbon cycle combined⁷, and extraction of CO₂ from the mantle has a critical influence on Earth's climate for timescales of 10⁸–10⁹ yr (ref. 7). The residence time of carbon in the mantle has been considered to exceed the age of the Earth^{7,8}, but may be lower if carbon is extracted efficiently by pervasive deep melting. The flux of CO₂ to ridges is controlled by the depth of initial melting, as are those of elements extracted by incipient melting, including rare gases, U, Th and K.

Seismological and geoelectrical anomalies apparently require incipient melting beneath mid-ocean ridges commencing at depths of 150–300 km (refs 4–6). This is far too deep for dry melting of peridotite⁹, leading to the suggestion that the observations are caused by the melting of pyroxenite pods⁴ or the melting of peridotite with small amounts of H₂O and/or CO₂ (refs 1, 2, 10). However, experimental studies show that the solidi of pyroxenites are not sufficiently deep^{11,12} and that the concentration of H₂O available beneath ridges is far too small to instigate melting at appropriate depths¹³. Deep melting incited by small amounts of CO₂ has been previously suggested^{14–17}, but experiments on simplified carbonated peridotite (CMAS, or CaO-MgO-Al₂O₃-SiO₂) imply that initial melting occurs at ~200 km (ref. 16) (earlier results¹⁸ are believed to be in error owing to inappropriate thermocouple geometries¹⁶). This depth is not sufficient to explain observed seismic reflections and anisotropy at 260–300 km (ref. 6). Owing to the fluxing effect of additional components such as Na and Fe (refs 9, 19), the solidus of natural carbonated peridotite may satisfy the geophysical constraints but to date has only been determined at relatively low pressures²⁰.

We conducted experiments with nominally anhydrous, carbonate-bearing fertile peridotite. Samples were contained in Pt-graphite double capsules and temperature was varied between 1,075 and

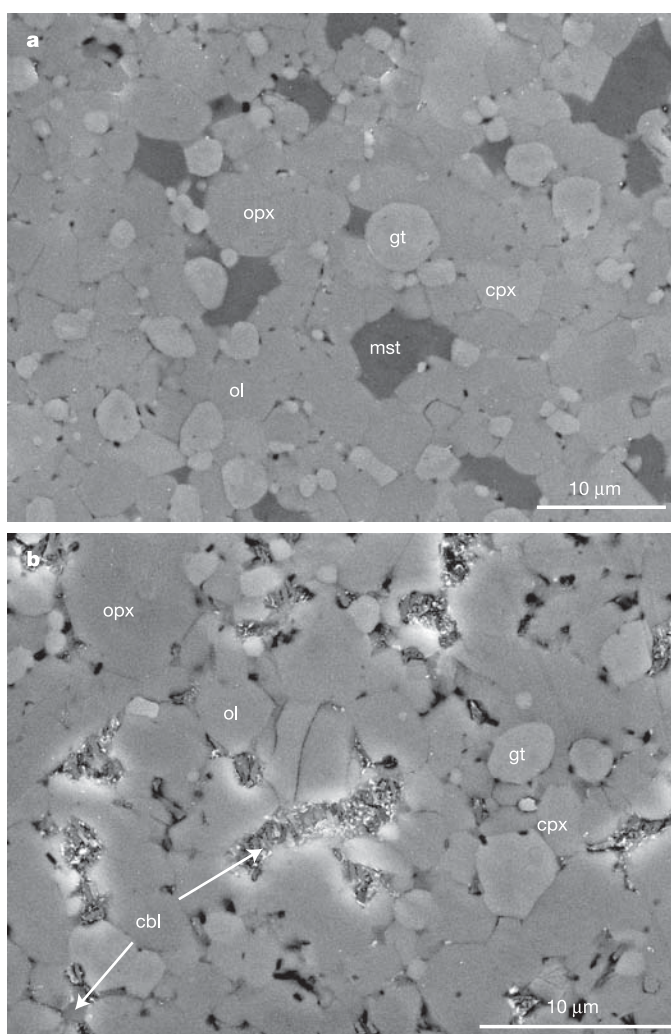


Figure 1 | Secondary electron images of typical run products, illustrating distinctions between melt-present and melt-absent conditions.

a, b, Discrete grains of magnesite are evident below the solidus (**a**; PERC at 6.6 GPa and 1,250 °C), but above the solidus quenched carbonate melts are interstitial to silicate grains (**b**; PERC at 6.6 GPa and 1,300 °C). Quenched carbonate melts are extremely fragile and are only partially preserved on polished surfaces (see Supplementary Methods). Abbreviations: ol, olivine; opx, orthopyroxene; cpx, clinopyroxene; gt, garnet; mst, magnesite; and cbl, carbonate melt.

¹Department of Geology and Geophysics, University of Minnesota, 310 Pillsbury Drive SE, Minneapolis, Minnesota 55455, USA.

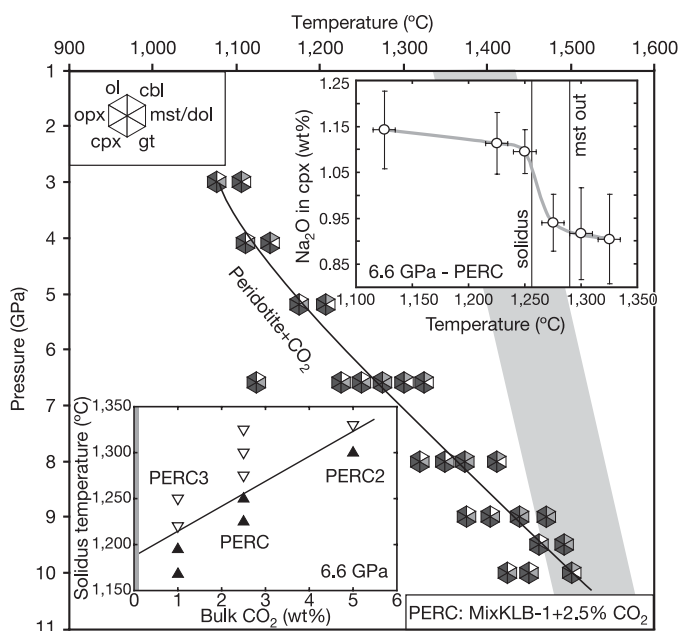


Figure 2 | Experimental constraints on the solidus of natural carbonated peridotite. Sectors of hexagons represent phases present (abbreviations as in Fig. 1, in addition to dol, dolomite solid solution). Silicates and carbonates (crystalline or molten) are indicated by dark and light grey shading. The grey band represents sub-ridge geotherms for a mantle potential temperature of 1,315–1,425 °C (refs 30 and 31). The top right inset shows the variation of Na₂O concentration in cpx with temperature for PERC composition at 6.6 GPa (the 1σ s.d. error bars displayed are based upon more than 10 replicate analyses). The sharp drop between 1,250 °C and 1,275 °C confirms the textural observation of carbonate melting. The bottom left inset demonstrates the effect of bulk CO₂ concentration on the solidus temperature of carbonated peridotite at 6.6 GPa. Open inverted triangles represent carbonate melt-present conditions while filled upright triangles are subsolidus experiments. The Na₂O/CO₂ (weight ratio) of PERC, PERC2 and PERC3 is 0.12, 0.06 and 0.30 respectively. The negative correlation between bulk Na₂O/CO₂ and solidus temperature demonstrates the solidus-lowering effect of sodium on carbonated peridotite. For CO₂ concentrations of 120–1,200 p.p.m. in the sub-ridge mantle source^{28,29} as shown by the vertical grey bar (see Supplementary Discussion), the solidus of natural mantle peridotite may be ~50 °C lower than that determined for PERC. If there is <10 p.p.m. CO₂, carbon may reside only in nominally carbon-free silicates at subsolidus conditions³² and the solidus will increase sharply, approaching the volatile-free peridotite solidus⁹ (~1,775 °C at 6.6 GPa) at a CO₂ concentration of zero.

1,500 °C, while a pressure of 3 to 10 GPa was applied using piston cylinder (for a pressure of 3 GPa) and Walker-style multi-anvil presses (for the pressure range 4 to 10 GPa). Peridotite with 2.5 wt% CO₂ (starting mix named 'PERC') was constructed from a fertile peridotite (MixKLB-1; Supplementary Table 1), and a mixture of natural and synthetic carbonates was added in proportion so as to maintain the ratio Ca:Mg:Fe:Na:K of the unmodified base peridotite (see Supplementary Methods and Supplementary Tables 1 and 2).

Olivine (ol), orthopyroxene (opx), clinopyroxene (cpx) and garnet (gt) were present in all of the experiments (Supplementary Table 3; Figs 1 and 2). The stable near-solidus crystalline carbonate consisted of dolomite (dol) at 3 GPa and magnesite (mst) from 4 to 10 GPa (Supplementary Table 3; Figs 1a and 2). Just above the solidus, carbonate melt (cbl) was present and crystalline carbonate disappeared within 20–50 °C of the solidus (Fig. 2, Supplementary Table 3). Textural criteria were used to identify the solidus: crystalline carbonate appeared as discrete grains while carbonate melt formed quenched mats in the interstices of silicate grains (Fig. 1). Solidi were also verified by tracking changes in the concentration of Na in cpx with temperature at pressures of 6.6, 8 and 9 GPa (Fig. 2). Sharp drops in Na concentrations in cpx coincide with textural evidence of melting (see Fig. 2), as a result of preferential partitioning of Na into carbonated melt²¹. The solidus of PERC increases from ≥1,075 °C at 3 GPa to 1,110–1,140 °C at 4.1 GPa as the stable carbonate at the solidus transforms from dolomite solid solution to magnesite solid solution. Above 4.1 GPa, the solidus of PERC magnesite lherzolite increases monotonically to approximately 1,500 °C at 10 GPa (Fig. 2). This solidus is consistent with that of natural carbonated peridotite within the pressure range 2 to 3.5 GPa (ref. 20).

Typical mantle peridotite has <1 wt% CO₂ (see Supplementary Information), but the detection of melt for such small CO₂ concentrations is not feasible. For natural bulk compositions of high thermodynamic variance, the extra CO₂ added to experiments to aid melt detection can bias solidus determinations. In particular, this can occur either if the starting composition possesses a bulk Na₂O/CO₂ fraction that differs from naturally occurring ratios, or if reactions between excess CO₂ and silicate minerals result in a near solidus crystalline carbonate with a Ca:Mg:Fe ratio that is different from that occurring in nature. Experiments with carbonated eclogite have demonstrated that the addition of a carbonate mixture with cation ratios similar to the base silicate has a minimal effect on the solidus¹⁹. However, excess CO₂ may still increase the solidus if the resulting carbonatite liquid has a cation ratio that is different from that of the bulk, or if the excess melt dilutes concentrations of

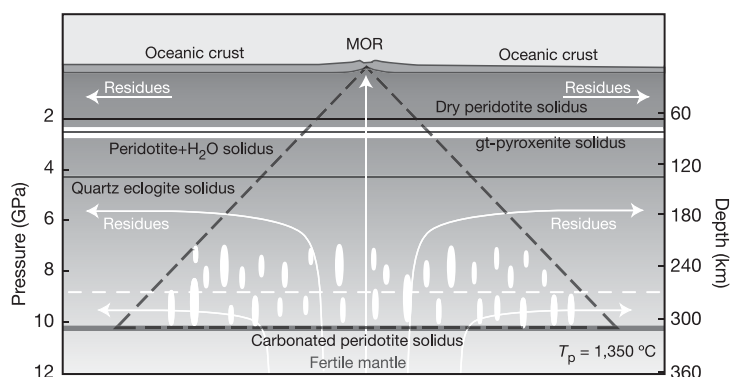


Figure 3 | Melting regime for passive upwelling beneath a mid-ocean ridge. Shown are the depths of solidi of different lithologies along a 1,350 °C (ref. 30) potential temperature (T_p) geotherm. The carbonated peridotite solidus corresponds to a concentration of 120–1,200 p.p.m. CO₂ in the mantle that is sufficiently oxidized to host carbon as carbonates. If carbon is present as diamond rather than carbonate, melting may initiate at shallower depths (see text). Also shown are solidi for damp (50–200 p.p.m. H₂O)

peridotite (white band)¹³, volatile-free peridotite⁹, garnet-pyroxenite¹² and quartz eclogite¹¹. Incipient melting owing to trace quantities of CO₂ begins 200–250 km deeper than melting of the volatile-free peridotite (at about 60 km)⁹. The horizontal white dashed line and vertically oriented ellipses respectively indicate the locations of the seismic reflector and vertical anisotropy detected beneath the East Pacific Rise⁶. MOR, mid-ocean ridge.

components such as Na_2O or K_2O that lower the solidus¹⁹. In order to test for the effect of excess CO_2 on the observed solidus of PERC, we performed additional experiments at 6.6 GPa with samples containing 1 wt% CO_2 (PERC3) and 5 wt% CO_2 (PERC2), but identical in all other respects (Supplementary Table 2). The solidus of PERC2 resides between 1,300 and 1,330 °C and that of PERC3 near 1,210 °C (Fig. 2 and Supplementary Fig. 1). Comparison of the 6.6 GPa solidus brackets for the three different bulk compositions (PERC, PERC2 and PERC3) indicates that the solidus of peridotite with very small carbonate content is similar to, but slightly (~ 50 °C) lower than, that measured for PERC (Fig. 2).

The solidus of naturally carbonated peridotite (PERC) is ~ 150 °C cooler than the solidus of $\text{CMAS} + \text{CO}_2$ (ref. 16) and at 10 GPa it is ~ 500 °C below the nominally volatile-free solidus⁹. The PERC solidus intersects the oceanic ridge geotherm (mantle potential temperature of 1,315–1,425 °C; refs 30 and 31) between 9.5 and 10.5 GPa, or at a depth of about 290–320 km (Fig. 2). Because natural mantle contains less CO_2 than PERC, this is an underestimate; the intersection of the ridge geotherm with the solidus of peridotite containing 120–1,200 p.p.m. of CO_2 (Supplementary Discussion) is 30 km deeper (Fig. 3). Such a small amount of CO_2 in the source will produce 0.03 to 0.3 wt% carbonatitic melt (these melts have ~ 40 wt% CO_2 ; refs 16, 19). The initiation of melting could be affected if carbon at depth is located in diamond rather than in carbonate²² (Fig. 4). If carbonate is not stable to the indicated carbonated peridotite solidus (~ 300 km), the onset of carbonatite stability will occur at a depth intermediate between ~ 300 km and the depth at which the transition from reduced to oxidized solid carbonate occurs^{22,23}. Estimates of oxygen fugacity in the sub-ridge mantle (see Fig. 4 legend) indicate that the diamond-carbonate transition may occur at 300 km or deeper, although shallower depths cannot be excluded (Fig. 4 and Supplementary Information).

Small amounts of carbonatite generated at a depth beneath ridges may be insufficient to affect directly observable geophysical properties in this region. Instead, we suggest that the anisotropy and seismic reflections⁶ are an indirect effect of melting and melt percolation. Here, incipient carbonate melt moves through the mantle beginning at a

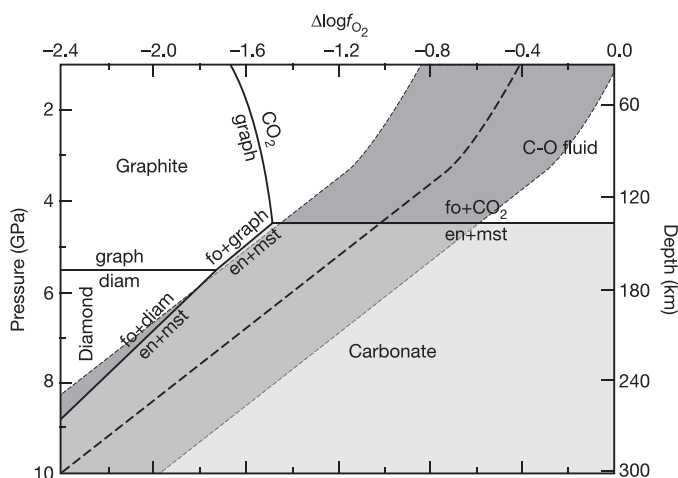


Figure 4 | Carbon storage and speciation along oceanic mantle adiabat. The dashed line indicates a change in oxygen fugacity (relative to the fayalite-magnetite-quartz buffer) with depth²², assuming constant $\text{Fe}^{3+}/\Sigma\text{Fe}$ and a mean estimate of the f_{O_2} of the mid-ocean ridge basalt (MORB)³³. The dark grey band encompasses the possible range of f_{O_2} based on the uncertainty in the $\text{Fe}^{3+}/\Sigma\text{Fe}$ estimate of MORB³³. The reaction labelled 'fo + diam = en + mst' shows the limit of stability of carbonate relative to diamond in peridotite. Abbreviations: fo, forsterite (ol); en, enstatite (opx); graph, graphite; mst, magnesite; and diam, diamond. This figure follows ref. 23.

depth of approximately 300 km and promotes extensive dissolution/precipitation of the olivine matrix (as is observed on experimental timescales²⁴), thus influencing deformation mechanisms and promoting grain coarsening and lattice preferred orientations²⁵. High mobility of carbonatite permits interconnectivity for melt fractions as low as 0.04 vol.% for 1 mm grains²⁶, which could potentially account for high electrical conductivity⁵. Finally, stiffening of thin intergranular carbonatitic melt films may inhibit upward percolation²⁷, thereby perhaps allowing for local enhancements of melt concentrations and the promotion of vertical melt channelling that is seismically detectable.

Small-degree carbonatite melts will be strongly enriched in, and residual peridotite markedly depleted of, highly incompatible elements, including those that produce heat (U, Th, K), rare gases and heavy alkalis and alkali-earth (Cs, Rb, Ba). Consequently, key radiogenic parent-daughter pairs (for example, U-Th-Pb-He and K-Ar, Rb-Sr) may be strongly fractionated in the depleted residual peridotite (Fig. 5). On the other hand, such melting could leave most of the H_2O in the residual peridotite (Fig. 5). We emphasize that the magnitude of these depletions and the direction of fractionation remain uncertain, given that partition coefficients for the appropriate phase compositions and conditions are poorly known. However, the depletions may be of great importance in the geochemical evolution of the mantle because the affected masses may be extremely large. For seafloor growth rates of $3 \text{ km}^2 \text{ yr}^{-1}$ and a mantle density of $3,300 \text{ kg m}^{-3}$, the flux of upper mantle passing upwards through the carbonated peridotite solidus depends on whether flow at that depth is focused beneath ridges—corner flow, as sketched in Fig. 3—or distributed throughout the mantle. In the case of the former, for melt extraction at 300 km, $3 \times 10^{18} \text{ g}$ of mantle will undergo incipient melting per year, of which $2.4 \times 10^{18} \text{ g}$ will not undergo major

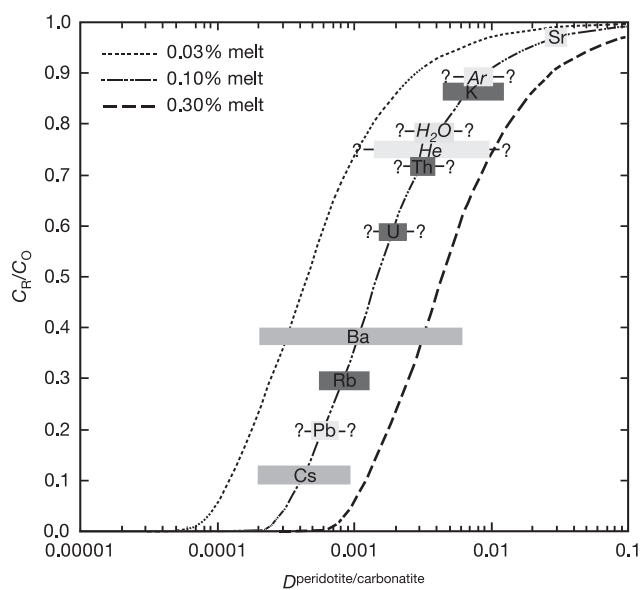


Figure 5 | Effect of removal of carbonatite melt on the concentration of incompatible trace elements in the residue (C_R) relative to the initial source composition (C_0) as a function of bulk partition coefficient $D^{\text{peridotite/carbonatite}}$. Depletions shown are for fractional melting ranging from 0.03% to 0.30%, corresponding to a concentration of 120 to 1,200 p.p.m. CO_2 in the source. Representative depletions for 0.1% melting are shown for key incompatible elements including radiogenic parents (darkest grey), daughters (lightest grey) and volatile elements (in italics), based on best estimates of $D^{\text{peridotite/carbonatite}}$ (Supplementary Information including Supplementary Table 4). Lengths of bounding boxes indicate plausible ranges $D^{\text{peridotite/carbonatite}}$ from these estimates. If the extraction of small-degree melts from peridotite matrices is incomplete, the calculated depletions represent maxima.

melting at depths of 60 km or shallower. In the latter case, the flux across the boundary must balance the creation and destruction of a 100-km-thick lithosphere, which amounts to $1 \times 10^{18} \text{ g yr}^{-1}$. Therefore, the mass undergoing incipient melting in 1 Gyr amounts to 25–75% of the mass of the mantle. The fate of such depleted regions is not well constrained. If parts of the incipiently depleted mantle are dragged downward with subducting slabs, they may form long-lived geochemical reservoirs. On the other hand, if most of the convecting mantle above about 300 km is remelted at ridges before being brought to the deeper mantle, its contribution to long-lived reservoirs may be small.

Extraction of small-degree melts from the mantle above 300 km implies residence times of 1–4 Gyr for carbon and other highly incompatible elements in the convecting mantle (unless there is a reservoir located in the deep mantle that is rich in carbon and trace elements). Such short residence times suggest that large fractions of mantle carbon are recycled rather than primordial. Carbonatite melts extracted at ~300 km provide a supply of carbon, rare gases and other highly incompatible elements to ridges. For a source containing 120–1,200 p.p.m. CO_2 , CO_2 fluxes are $(0.12\text{--}3.40) \times 10^{15} \text{ g yr}^{-1}$, which matches or exceeds direct flux estimates at ridges $((0.10\text{--}0.66) \times 10^{15} \text{ g yr}^{-1}$; refs 28 and 29). Efficient extraction of carbon and highly incompatible trace elements from such large source volumes may reduce the absolute concentrations in the sub-ridge mantle that are needed to account for the observed fluxes and concentrations at ridges. However, not all of the small-degree melt formed at large depth may be extracted from its source and not all of the extracted melt may reach ridges; some could instead be implanted into the oceanic lithosphere, thus providing a widespread source for metasomatic fluids that are rich in incompatible elements.

Received 18 July 2005; accepted 26 January 2006.

- Plank, T. & Langmuir, C. H. Effects of melting regime on the composition of the oceanic crust. *J. Geophys. Res.* **97**, 19749–19770 (1992).
- Galer, S. J. G. & O'Nions, R. K. Magmagenesis and the mapping of chemical and isotopic variations in the mantle. *Chem. Geol.* **56**, 45–61 (1986).
- Karato, S.-I. & Jung, H. Water, partial melting and the origin of the seismic low velocity and high attenuation zone in the upper mantle. *Earth Planet. Sci. Lett.* **157**, 193–207 (1998).
- The MELT Seismic Team. Imaging the deep seismic structure beneath a mid-ocean ridge: the MELT experiment. *Science* **280**, 1215–1218 (1998).
- Evans, R. L. *et al.* Asymmetric electrical structure in the mantle beneath East Pacific Rise at 17°S. *Science* **286**, 752–756 (1999).
- Gu, Y. J., Lerner-Lam, A. L., Dziewonski, A. M. & Ekström, G. Deep structure and seismic anisotropy beneath the East Pacific Rise. *Earth Planet. Sci. Lett.* **232**, 259–272 (2005).
- Sleep, N. H. & Zahnle, K. Carbon dioxide cycling and implications for climate on ancient Earth. *J. Geophys. Res.* **106**, 1373–1399 (2001).
- Zhang, Y. & Zindler, A. Distribution and evolution of carbon and nitrogen in Earth. *Earth Planet. Sci. Lett.* **117**, 331–345 (1993).
- Hirschmann, M. M. The mantle solidus: experimental constraints and the effect of peridotite composition. *Geochem. Geophys. Geosyst.* **1**, 2000GC000070 (2000).
- McKenzie, D. The extraction of magma from the crust and mantle. *Earth Planet. Sci. Lett.* **74**, 81–91 (1985).
- Yasuda, A., Fujii, T. & Kurita, K. Melting phase relations of anhydrous mid-ocean ridge basalt from 3 to 20 GPa: implications for the behavior of subducted oceanic crust in the mantle. *J. Geophys. Res.* **99**, 9401–9414 (1994).
- Kogiso, T., Hirschmann, M. M. & Frost, D. J. High-pressure melting of garnet-pyroxenite: possible mafic lithologies in the source of ocean island basalts. *Earth Planet. Sci. Lett.* **216**, 603–617 (2003).
- Aubaud, C., Hauri, E. H. & Hirschmann, M. M. Hydrogen partition coefficients between nominally anhydrous minerals and basaltic melts. *Geophys. Res. Lett.* **31**, L20611, doi:10.1029/2004GL021341 (2004).
- Wyllie, P. J. & Huang, W.-L. Influence of mantle CO_2 in the generation of carbonatites and kimberlites. *Nature* **257**, 297–299 (1975).
- Eggler, D. H. Does CO_2 cause partial melting in the low-velocity layer of the mantle? *Geology* **4**, 69–72 (1976).
- Dalton, J. A. & Presnall, D. C. Carbonatitic melts along the solidus of model lherzolite in the system $\text{CaO-MgO-Al}_2\text{O}_3\text{-SiO}_2\text{-CO}_2$ from 3 to 7 GPa. *Contrib. Mineral. Petrol.* **131**, 123–135 (1998).
- Presnall, D. C. & Gudfinsson, G. H. in *Plates, Plumes, and Paradigms* (eds Foulger, G. R., Natland, J. H., Presnall, D. C. & Anderson, D. L.) 207–216 (Special Paper 388, Geological Society of America, Boulder, 2005).
- Canil, D. & Scarfe, C. M. Phase relations in peridotite+ CO_2 systems to 12 GPa: implications for the origin of kimberlite and carbonate stability in the Earth's upper mantle. *J. Geophys. Res.* **95**, 15805–15816 (1990).
- Dasgupta, R., Hirschmann, M. M. & Dellas, N. The effect of bulk composition on the solidus of carbonated eclogite from partial melting experiments at 3 GPa. *Contrib. Mineral. Petrol.* **149**, 288–305 (2005).
- Falloon, T. J. & Green, D. H. The solidus of carbonated fertile peridotite. *Earth Planet. Sci. Lett.* **94**, 364–370 (1989).
- Wallace, M. E. & Green, D. H. An experimental determination of primary carbonatite magma composition. *Nature* **335**, 343–346 (1988).
- Wood, B. J., Pawley, A. & Frost, D. R. Water and carbon in the Earth's mantle. *Phil. Trans. R. Soc. Lond.* **354**, 1495–1511 (1996).
- Frost, D. J. & Wood, B. J. Experimental measurements of the fugacity of CO_2 and graphite/diamond stability from 35 to 77 kbar at 925 to 1650 °C. *Geochim. Cosmochim. Acta* **61**, 1565–1574 (1997).
- Hammouda, T. & Laporte, D. Ultrafast mantle impregnation by carbonatite melts. *Geology* **28**, 283–285 (2000).
- Holtzman, B. K. *et al.* Melt segregation and strain partitioning: Implications for seismic anisotropy and mantle flow. *Science* **301**, 1227–1230 (2003).
- Minarik, W. G. & Watson, E. B. Interconnectivity of carbonate melt at low melt fraction. *Earth Planet. Sci. Lett.* **133**, 423–437 (1995).
- Rabinowicz, M., Ricard, Y. & Grégoire, M. Compaction in a mantle with a very small melt concentration: implications for the generation of carbonatitic and carbonate-bearing high alkaline mafic melt impregnations. *Earth Planet. Sci. Lett.* **203**, 205–220 (2002).
- Javoy, M. & Pineau, F. The volatiles record of a 'popping' rock from the Mid-Atlantic Ridge at 14°N: chemical and isotopic composition of gas trapped in the vesicles. *Earth Planet. Sci. Lett.* **107**, 598–611 (1991).
- Marty, B. & Tolstikhin, I. N. CO_2 fluxes from mid-ocean ridges, arcs, and plumes. *Chem. Geol.* **145**, 233–248 (1998).
- Ita, J. & Stixrude, L. Petrology, elasticity, and composition of the mantle transition zone. *J. Geophys. Res.* **97**, 6849–6866 (1992).
- McKenzie, D., Jackson, J. & Priestley, K. Thermal structure of oceanic and continental lithosphere. *Earth Planet. Sci. Lett.* **233**, 337–349 (2005).
- Keppler, H., Wiedenbeck, M. & Shcheka, S. S. Carbon solubility in olivine and the mode of carbon storage in the Earth's mantle. *Nature* **424**, 414–416 (2003).
- Bézos, A. & Humler, E. The $\text{Fe}^{3+}/\Sigma\text{Fe}$ ratios of MORB glasses and their implications for mantle melting. *Geochim. Cosmochim. Acta* **69**, 711–725 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. C. Withers and C. Aubaud for comments on the manuscript, P. Asimow for conversations and N. Smith for help with the piston cylinder experiments. This work is supported by NSF.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.D. (dasg0007@umn.edu).

Altruism through beard chromodynamics

Vincent A. A. Jansen¹ & Minus van Baalen²

The evolution of altruism, a behaviour that benefits others at one's own fitness expense, poses a darwinian paradox. The paradox is resolved if many interactions are with related individuals so that the benefits of altruism are reaped by copies of the altruistic gene in other individuals¹, a mechanism called kin selection². However, recognition of altruists could provide an alternative route towards the evolution of altruism^{1,3–5}. Arguably the simplest recognition system is a conspicuous, heritable tag, such as a green beard^{1,3}. Despite the fact that such genes have been reported^{6–8}, the 'green beard effect'³ has often been dismissed because it is unlikely that a single gene can code for altruism and a recognizable tag^{1,3,9}. Here we model the green beard effect and find that if recognition and altruism are always inherited together, the dynamics are highly unstable, leading to the loss of altruism. In contrast, if the effect is caused by loosely coupled separate genes, altruism is facilitated through beard chromodynamics in which many beard colours co-occur. This allows altruism to persist even in weakly structured populations and implies that the green beard effect, in the form of a fluid association of altruistic traits with a recognition tag, can be much more prevalent than hitherto assumed.

If every individual were to behave altruistically the population as a whole would do well. That altruism nevertheless does not readily evolve is illustrated by evolution of cooperation in the prisoner's dilemma game¹⁰. In this game a player can either help another player by cooperating (playing C), or not help by defecting (playing D). Because the payoff of cooperation is always less than that of defection, cooperation is costly and is thus an act of altruism. As defectors always do better than cooperators in the same situation, cooperation cannot evolve in large, well-mixed populations in which different players are encountered every round, even though the highest average payoff is realized in a population in which all players cooperate. Cooperation can evolve in sufficiently viscous populations where patterns of relatedness create a population structure that allows kin selection to operate¹¹, if the benefits of cooperation outweigh the effects of kin competition^{12,13}.

The evolution of altruism is obviously facilitated by mechanisms that allow discrimination against defectors^{1,14}. One such mechanism is the green beard effect in which altruists can recognize each other using a conspicuous tag or signal^{1,3}. In Dawkins³ formulation of the green beard effect this is achieved through a single gene causing both altruistic behaviour and recognition. This tight coupling has been considered a crucial characteristic for the green beard effect to work⁴ because if the genes for tag and altruistic trait were loosely coupled then not only altruists can have coloured beards, but also non-altruists would acquire them. Such individuals would receive the benefits of altruistic behaviour without having to pay the cost: they cheat on the interaction, thus potentially preventing the evolution of altruism. However, because a gene that causes both traits is considered to be too complex to be likely^{1,3,9} the green beard effect has often been considered implausible.

A number of observations have suggested that the green beard

effect actually does exist in nature^{6–8,15}, and results from simulation studies suggest that altruism can be maintained through the co-existence of a small number of beard colours^{16,17}. Here, we will investigate theoretically if and when the green beard effect can operate. We will do this first for tightly coupled genes for tag and trait. Then, we will explore the consequences of an idea originally suggested by ref. 18: that there exist separate genes, one for beard colour, which facilitates recognition, and one for being altruistic or not. These genes are loosely coupled in that they can be inherited separately and thus can give rise to new tag–trait combinations.

The model that we use for the dynamics of beard colour polymorphism (beard chromodynamics) is based on the prisoner's dilemma game in a spatial setting^{12,17,19}, with one extra twist. As in refs 16 and 17, we assume that all individuals have a recognizable tag in the form of a coloured beard and that altruistic actions are only towards individuals with the same beard colour. Each individual plays all its neighbours and receives a score according to the payoff matrix in Table 1. Players put offspring in empty neighbouring sites with a probability proportional to their score. Apart from rare mutations, new tag–trait combinations arise because reproducing individuals have a certain probability to mate with a neighbouring individual, swapping part of their genomes in the process. We carried out explicit simulations (see Supplementary Information for details) of full interaction networks in which we varied the connectedness and topology of the interaction network. We also performed a more in-depth analysis by considering the set of replicator equations²⁰ that results when a simplifying assumption is made (see Methods and Supplementary Information).

If only a single beard colour is present, no discrimination occurs and the model describes blind kin selection. This can maintain altruism only if the population is sufficiently viscous and the scales of cooperation and competition are sufficiently different^{16,13}. Both simulations and mathematical analysis of our model show that the green beard effect does not enhance the possibilities for altruism if tag and trait are always inherited together, because this leads to highly unstable dynamics. To understand why, consider a cooperator with a rare beard colour in a population dominated by other beard colours.

Table 1 | Payoff matrix for the multi-beard prisoner's dilemma

Payoff	C _i	D _i	C _j	D _j
C _i	R	S	P	P
D _i	T	P	P	P
C _j	P	P	R	S
D _j	P	P	T	P

The payoff of a player using a strategy in the first column against a player using a strategy in the first row with $j \neq i$. The strategies are to cooperate (C) and to defect (D); the subscript indicates beard colour. We assumed $T > R > P > S$ and that $T + S > P + R$. For these parameters, the highest payoff is received if one has an opponent who actually cooperates, irrespective of one's own strategy. Because the payoff of a cooperator never exceeds that of a defector, irrespective of the opponent, cooperation is costly for the perpetrator and therefore is an altruistic strategy.

¹School of Biological Sciences, Royal Holloway, University of London, Egham, Surrey TW20 0EX, UK. ²Laboratoire d'Ecologie ENS/UPMC/CNRS, UMR 7625, Université Pierre et Marie Curie, Bâtiment A, 7ème Etage, CC 237, 7 quai St Bernard, 75252 Paris Cedex 05, France.

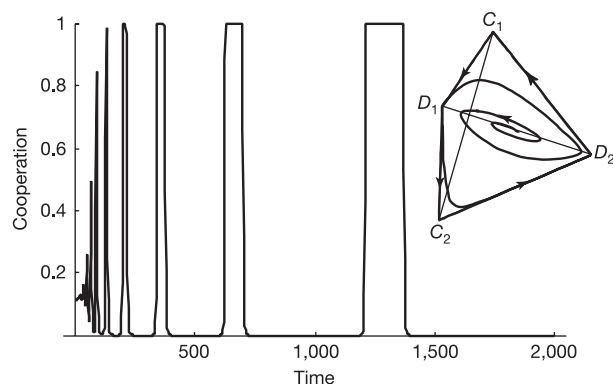


Figure 1 | The level of cooperation in the population in the approach to the heteroclinic cycle. The system, specified by equation (1), spends an increasing amount of time in equilibria with a single beard colour. Inset: the dynamics on a three-dimensional simplex. Parameters: $T = 5$, $R = 3$, $P = 1$, $S = 0$, $h = 1$.

This cooperator will behave as a defector when meeting individuals with different beard colours. It can therefore invade the population if the average cooperation level (and hence the average payoff) is below a threshold level and will eventually come to dominate this population, erasing any beard colour diversity that may have existed up to that point. Once the new colour dominates, defectors carrying beards in this new colour can successfully invade. Hence, the initial fitness advantage of the new beard is lost, with the sole result that the population has changed to a new, single colour. The (re-)invasion of cooperators in rare beard colours, followed by the emergence of defectors in this colour repeats indefinitely. If beard colour and strategy are always inherited together this scenario corresponds to highly unstable dynamics that result in the rapid loss of beard colours. This is confirmed by our analysis (see Supplementary Information), which reveals that if tag and trait are tightly coupled the dynamics are dominated by an attracting heteroclinic cycle on which the population is monochrome for most of the time (Figs 1 and 2).

If, however, tag and trait are coded by separate, loosely linked genes a different pattern emerges. Loose coupling results in dynamics that are less unstable and in which cooperation arises through the dynamic coexistence of different beard colours (Fig. 2). The reason for this difference is that loose coupling prevents a single beard colour from dominating the dynamics. Whereas tightly coupled genes create dynamics that go through cycles with ever deeper troughs in which eventually fixation occurs, loose coupling continuously generates new tag–trait combinations that prevent fixation and stabilizes the dynamics. In our spatial simulations this boom–bust scenario can be observed through clusters of cooperators with same-coloured beards in an environment otherwise dominated by defectors. These clusters increase in size over time until a defector with the same beard colour appears in the vicinity of the cluster. Once this happens the cluster is taken over by defectors, until a cooperator in a novel beard colour appears. This results in a shifting mosaic of beard colours (Fig. 3). The tighter the coupling, the smaller the chance of a new type appearing and hence the larger the size of these clusters and the more unstable the dynamics are. The crucial mechanism that stabilizes the dynamics is the regular local appearance of novel tag–trait combinations. In our model such new combinations are created through recombination, which we consider the most likely mechanism to operate in natural populations. However, other mechanisms that can introduce new heritable tag–trait combinations in local clusters, such as high levels of mutation, infrequent long distance dispersal or gene flow, can accomplish the same stabilizing effect^{16,17}.

Under recombination, successive invasions of new beard colours tend to increase beard-colour diversity up to a point where no new

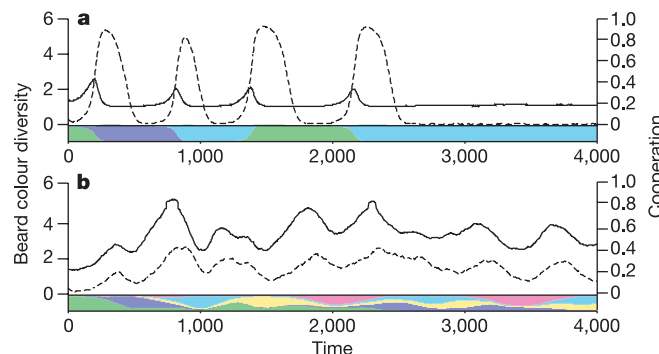


Figure 2 | Evolution of cooperation and beard colour diversity. The dynamics of the overall level of cooperation (dashed line) and beard colour diversity (solid line) in the simulation model. Beard colour diversity is measured as the exponential of the Shannon index ($\exp[-\sum_i (C_i + D_i) \ln(C_i + D_i)]$), which returns the number of beard colours when they are present in equal densities, and equilibrates well below the maximum value of 6. The simulations were run on a random network of 10,000 sites where every site has eight connections to other sites; all other parameter values are as in Fig. 1 with $h = 1$ (a) and $h = 0.9025$ (b) (corresponding to $\rho = 0.95$ in the simulation model), and the mutation rate was 10^{-4} per locus. The lattices were seeded with a small proportion of blue-bearded cooperators in a population of random genotypes with a strong bias towards green-bearded defectors. In this simulation cooperation cannot be maintained through blind kin selection alone.

beard colours can invade and diversity saturates (Fig. 2). Our analysis reveals how this diversity is regulated: cooperators will, on the whole, encounter fewer defectors with a similar beard colour and thus be exploited less if the diversity in beard colours is high. Therefore, the average payoff increases with the number of beard colours that are established in the population. Because all mutant beard colours have a constant fitness when rare (independent, in particular, of beard colour diversity), it becomes increasingly difficult for new colours to establish themselves in the population when the beard colour diversity increases (see Methods and Supplementary Information). Beard colour diversity is thus regulated at a definite level (Figs 2 and 4). Our analysis confirms that loose coupling is necessary for the evolution of tag-based cooperation. Nevertheless, the coupling should not be too loose. Coupling that is too loose causes the correlation between tag and trait to be too weak for the tag to serve as a proxy for the trait, whereas a coupling that is too tight means not only that clusters of cooperators will be homogeneous for beard colour, but also that the consequences will be dramatic when neighbouring cheats eventually acquire the same beard colour.

The final beard colour diversity depends on viscosity. An increase in connectedness leads to a decrease in the average cooperation level and the average payoff, if the number of beard colours is kept constant. However, a lower average payoff allows new beard colours to become established, resulting in an increase in beard colour diversity, which, in turn, counteracts the decrease in the level of cooperation (Fig. 4). This mechanism causes the number of beard colours to be negatively correlated with viscosity: the less the population is structured, the higher the beard colour diversity. Importantly, this mechanism maintains altruism in populations in which blind kin selection alone cannot.

Our model assumes that the green beard effect works through alleles that cause both the tag and recognition of that same tag. The assumption of a single recognition allele is justified if the allele functionally combines the tag and recognition functions, as is conceivable for homophilic cell surface adhesion proteins^{6,8,15} or if recognition is self-referent and works through comparing another individual's tag to one's own. Then, a single mutation can change both one's tag and recognition of the tag. Alternatively, recognition could be based on two different alleles, one for the tag and one for

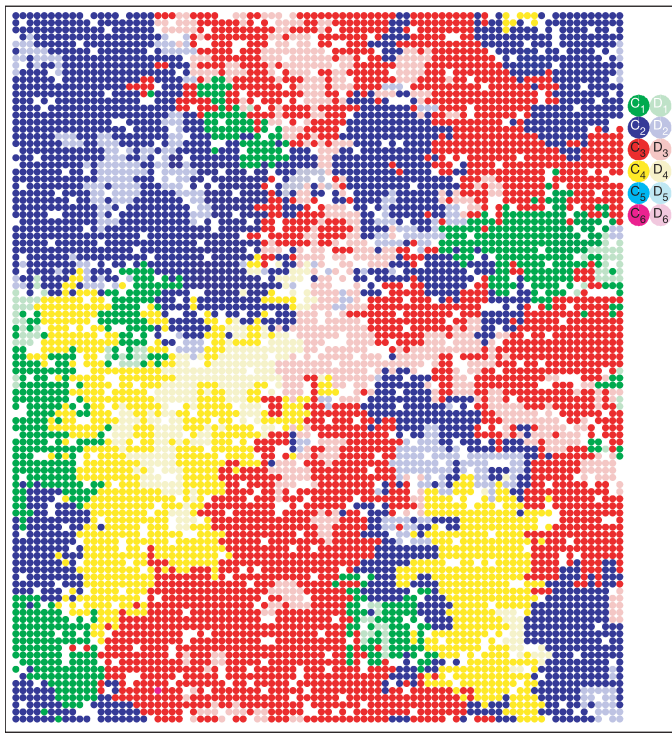


Figure 3 | An example of a snapshot of spatial beard chromodynamics. Snapshot taken at $t = 4,000$ on a square lattice (with four neighbours per site). Altruistic individuals are indicated by dark colours; defectors by light colours. Parameters match those of Fig. 2 but with $h = 0.81$ (corresponding to $\rho = 0.90$ in the simulation model).

recognition. Even though this case is not strictly covered by our model, our qualitative results should carry over. Because a mismatch between tag and recognition function is neutral in defectors (relative to the fitness of other defectors with a similar tag), diversity in these alleles can build up in the defector population, providing the potential for the generation, through subsequent mutation on the second allele, of new matching sets of tag and recognition alleles. Even if this process is potentially slow, once it has created a set of matching alleles these will be maintained through selection, as described in our model. Moreover, because selection acts against mismatches in cooperators, one could conjecture that a tight coupling between tag and recognition alleles, as assumed in our model and found in nature⁷, naturally arises.

Our results imply that the scope for green beard genes is much wider than often assumed. This is for a number of reasons. First, altruism can be maintained without all the functions for tag, recognition and altruism having to reside in a single locus: loose coupling between a recognition allele and altruistic trait suffices. Second, our results suggest that rather than there being a single green beard gene in a population, one can expect to find a diversity of such genes, especially if the population is weakly structured. A possible reason that so few coloured beards have been reported is the concentration of research on highly structured populations in which the diversity of beard colours is predicted to be low. Our analysis leads to the testable hypothesis that diversity in recognition tags inversely correlates with average relatedness. This suggests that relatively easily observed tag diversity can serve as an indicator for the nature of the underlying social interactions. Third, to detect the green beard effect one should look for cases where tag and trait can dynamically associate^{8,16}. A tag that functions as a green beard in one instance need not be associated with altruism in another population or at another instance in time, which obviously has consequences for our capacity to detect green beards.

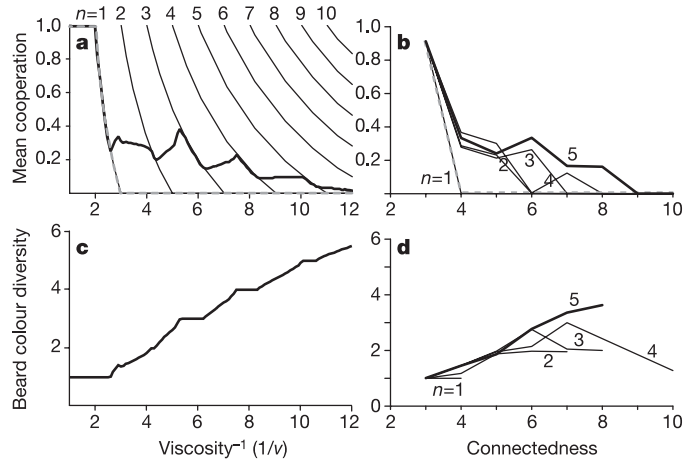


Figure 4 | Cooperation and diversity versus viscosity. **a**, The mean cooperation in the model (equation (1)) is found by letting the beard colour diversity saturate (thick line). The thin lines give the mean cooperation on the symmetrical equilibria for different numbers of beard colours; the grey dashed line is the result for a single beard colour, corresponding to blind kin selection. **b**, As for **a** but for the simulation model. The different curves show mean values over $t \in [3,000, 4,000]$ in simulations on random networks with different average connectivity. The curves differ in the maximum number of beard colours used in the simulation. **c**, The beard colour diversity corresponding to **a**. **d**, The beard colour diversity corresponding to **b**. Parameters are as in Fig. 3.

METHODS

The mathematical model that we use is based on the simulation model (see Supplementary Information) and is inspired by the replicator equation²⁰. To capture the effects of viscosity we assumed that an individual encounters with a probability v an individual identical at both loci, and with a probability $1 - v$ a random individual from the population. This probability is approximately inversely proportional to the number of neighbours¹¹. This results in the following payoffs ϕ_x for phenotype x :

$$\phi_{C_i} = vR + (1 - v)(RC_i + SD_i + (1 - C_i - D_i)P)$$

$$\phi_{D_i} = vP + (1 - v)(TC_i + (1 - C_i)P)$$

The parameters R , P , T and S specify the payoff as defined in Table 1, and C_i (or D_i) is the fraction of the population that has phenotype C_i (or D_i). We also assumed that with a probability, ρ , a gene is inherited from a neighbouring individual. For reasons of simplicity we ignored in the mathematical model the small probability that an individual recombines both tag and trait, and therefore the probability of having the same genotype as the parent is $h = 1 - 2\rho$. Because a neighbouring individual has the same genotype with probability v the effective rate of recombination is $\rho' = (1 - v)\rho$ and players give rise to an exact copy of themselves with probability $h' = h + v(1 - h)$. By changing the parameter h , and thus ρ , we can change the level of linkage between beard colour and altruistic trait. This leads to the system of replicator equations:

$$\begin{aligned} \dot{C}_i &= (h' \phi_{C_i} - \Phi)C_i + \rho' \Phi_i \sum_{j=1}^n C_j + \rho' \Phi_C (C_i + D_i) \\ \dot{D}_i &= (h' \phi_{D_i} - \Phi)D_i + \rho' \Phi_i \sum_{j=1}^n D_j + \rho' \Phi_D (C_i + D_i) \end{aligned} \quad (1)$$

where $\Phi = \sum_{j=1}^n \phi_{C_j} C_j + \phi_{D_j} D_j$ represents the average fitness in the population, $\Phi_i = \phi_{C_i} C_i + \phi_{D_i} D_i$ represents the average fitness of individuals with beard colour i , $\Phi_C = \sum_{j=1}^n \phi_{C_j} C_j$ represents the average fitness of cooperators, and $\Phi_D = \sum_{j=1}^n \phi_{D_j} D_j$ represents the average fitness of defectors. This formalism assumes that the death rate is equal to the average fitness so that the total population size remains constant. We implemented mutation by infrequently and randomly changing tag or traits. If the total density of a certain beard colour dropped below 0.0005 we removed this beard colour and normalized the densities.

Because of symmetry between beard colours there exist equilibria in which all beard colours have equal densities. By putting the left-hand sides of equation (1)

to zero, by denoting the equilibrium densities and payoffs by bars and $\bar{C}_i = \bar{C}(n)$, $\bar{D}_i = \bar{D}(n)$ we find by elimination that $\bar{\phi}_{C_i} = \bar{\phi}_{D_i} = \bar{\Phi}$; that is, at equilibrium the average payoff of a cooperator and a defector is equal. Using this and $\bar{C}(n) + \bar{D}(n) = 1$ we find that

$$\bar{C}(n) = \frac{\nu(R - P) - \frac{1-\nu}{n}(P - S)}{(1 - \nu)(T + S - R - P)}.$$

Therefore, because $R > P$, $P > S$ and $T + S > R + P$, the equilibrium density of cooperators with a specific beard colour increases with the number of beard colours, as does the total amount of cooperation, $n\bar{C}(n)$. The average payoff at equilibrium, $\bar{\Phi} = P + (1 - \nu)(T - P)\bar{C}(n)$, increases with the number of beard colours because $T > P$.

The simulation model differs from the replicator model (equation (1)) in that, for the replicator model the density-dependent regulation acts globally, whereas in the simulation all density dependence is local. Especially if the effective scale on which local regulation operates is of a similar order as the scale over which the altruistic interactions take place, then local regulation can reduce the possibility of altruism¹³. The fact that altruism can be maintained in our simulation model through blind kin selection—provided the viscosity is sufficiently high—demonstrates that this is not an overriding effect. The fact that the different models give qualitatively similar results illustrates that the maintenance of altruism through beard chromodynamics does not critically depend on this aspect.

Received 7 October; accepted 28 October 2005.

- Hamilton, W. D. The genetic evolution of social behaviour, II. *J. Theor. Biol.* **7**, 17–52 (1964).
- Maynard Smith, J. Group selection and kin selection. *Nature* **201**, 1145–1147 (1964).
- Dawkins, R. *The Selfish Gene* (Oxford Univ. Press, Oxford, 1976).
- Dawkins, R. *The Extended Phenotype* (Oxford Univ. Press, Oxford, 1982).
- Riolo, R., Cohen, M. D. & Axelrod, R. Evolution of cooperation without reciprocity. *Nature* **414**, 441–443 (2001).
- Queller, D. C., Ponte, E., Bozzaro, S. & Strassmann, J. E. Single-gene greenbeard effects in the social amoeba *Dictyostelium discoideum*. *Science* **299**, 105–106 (2003).
- Keller, L. & Ross, K. G. Selfish genes: a green beard in the red fire ant. *Nature* **394**, 573–575 (1998).
- Summers, K. & Crespi, B. Cadherins in maternal-foetal interactions: red queen with a green beard? *Proc. R. Soc. Lond. B* **272**, 643–649 (2005).
- Blaustein, A. R. Kin recognition mechanisms—phenotypic matching or recognition alleles. *Am. Nat.* **121**, 749–754 (1983).
- Trivers, R. The evolution of reciprocal altruism. *Q. Rev. Biol.* **46**, 35–57 (1971).
- Van Baalen, M. & Rand, D. The unit of selection in viscous populations and the evolution of altruism. *J. Theor. Biol.* **193**, 631–648 (1998).
- Nowak, M. A. & May, R. M. Evolutionary games and spatial chaos. *Nature* **355**, 826–829 (1992).
- West, S. A., Pen, I. & Griffin, A. S. Conflict and cooperation—Cooperation and competition between relatives. *Science* **296**, 72–75 (2000).
- Hamilton, W. D. in *Kin Recognition in Animals* (eds Fletcher, D. J. C. & Michener, C. D.) 417–438 (Wiley, Chichester, 1987).
- Haig, D. Gestational drive and the green-bearded placenta. *Proc. Natl Acad. Sci. USA* **93**, 6547–6551 (1996).
- Van Baalen, M. & Jansen, V. A. A. Common language or Tower of Babel? On the evolutionary dynamics of signals and their meanings. *Proc. R. Soc. Lond. B* **270**, 69–76 (2003).
- Axelrod, R., Hammond, R. A. & Grafen, A. Altruism via kin-selection strategies that rely on arbitrary tags with which they coevolve. *Evolution* **58**, 1833–1838 (2004).
- Haig, D. in *Behavioural Ecology* 4th edn (eds Krebs, J. R. & Davies, N. B.) 284–306 (Cambridge Univ. Press, Cambridge, 1997).
- Axelrod, R. *The Evolution of Cooperation* (Basic Books, New York, 1984).
- Hofbauer, J. & Sigmund, K. *Evolutionary Games and Population Dynamics* (Cambridge Univ. Press, Cambridge, 1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank P. Credland for his encouragement and comments on the manuscript.

Author Contributions Both authors contributed equally to this paper. V.A.A.J. formulated and analysed the mathematical model; M.v.B. formulated and analysed the simulation model.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to V.A.A.J. (vincent.jansen@rhul.ac.uk) or M.v.B. (minus.van.baaen@ens.fr).

Chance and necessity in the evolution of minimal metabolic networks

Csaba Pál^{1,2*}, Balázs Papp^{3*}, Martin J. Lercher^{1,4}, Péter Csermely⁵, Stephen G. Oliver³ & Laurence D. Hurst⁴

It is possible to infer aspects of an organism's lifestyle from its gene content¹. Can the reverse also be done? Here we consider this issue by modelling evolution of the reduced genomes of endosymbiotic bacteria. The diversity of gene content in these bacteria may reflect both variation in selective forces and contingency-dependent loss of alternative pathways. Using an *in silico* representation of the metabolic network of *Escherichia coli*, we examine the role of contingency by repeatedly simulating the successive loss of genes while controlling for the environment. The minimal networks that result are variable in both gene content and number. Partially different metabolisms can thus evolve owing to contingency alone. The simulation outcomes do preserve a core metabolism, however, which is over-represented in strict intracellular bacteria. Moreover, differences between minimal networks based on lifestyle are predictable: by simulating their respective environmental conditions, we can model evolution of the gene content in *Buchnera aphidicola* and *Wigglesworthia glossinidia* with over 80% accuracy. We conclude that, at least for the particular cases considered here, gene content of an organism can be predicted with knowledge of its distant ancestors and its current lifestyle.

Naturally evolved, nearly minimal gene sets in closely related intracellular symbionts contain substantial differences². The diversity of these evolved minimal gene sets may be the product of three fundamental processes: differences in initial genetic makeup; variation in selective forces within host cells; and differences in the order of gene deletions, resulting in a choice between alternative cellular pathways². By modelling the reductive evolution of a detailed metabolic network, we first explore the evolutionary significance of the last of these alternatives.

Using the metabolic network of *Escherichia coli* K12 (ref. 3) as our model system has several advantages. First, the best evidence for the presence of alternative pathways within and across species comes from studies of metabolic networks⁴. Second, flux balance analysis provides a rigorous modelling framework for studying the impact of gene deletions^{4,5}; the method relies on optimizing the steady-state use of the metabolic network to produce biomass components. Third, not only is the metabolic network of *E. coli* K12 one of the best studied cellular subsystems, but this organism is also a close relative of several endosymbiotic organisms⁶, including *Buchnera aphidicola* and *Wigglesworthia glossinidia*. Cellular domestication has resulted in the elimination of 70–75% of the ancestral genome in these latter organisms⁷.

The previously reconstructed metabolic network of *E. coli*³ consists of 904 genes and 931 unique biochemical reactions, and incorporates external nutrients and the corresponding transport processes. The composition of a 'minimal reaction set' has been previously shown to

depend strongly on the given environmental conditions⁸. Gradual evolution towards minimal genomes and the role of chance in this process, however, have remained unexplored. The smallest sets of genes that are compatible with cellular life will relate to the most favourable conditions, in which most nutrients are available from the environment. This situation is approximated by organisms with a strict intracellular lifestyle, where the host provides most of their nutrients². Accordingly, we first characterized the simulated evolution of the network under nutrient-rich conditions (Supplementary Tables 1–3).

To explore systematically the combinatorial set of minimal metabolic reaction sets, we elaborated a simple algorithm for simulating gradual loss of metabolic enzymes. We remove a randomly chosen gene from the network and calculate the impact of this deletion on the production rate of biomass components (a proxy for fitness). If this rate is nearly unaffected, the deletion is assumed to be viable and the enzyme is considered to be permanently lost; otherwise, the gene is restored to the network. This procedure is repeated until no further enzymes can be deleted; that is, all remaining genes are essential for survival of the cell. This simulation was repeated 500 times, with each run providing an independent evolutionary outcome.

The resulting networks share on average 77% of their reactions, whereas only 25% would be shared by randomly deleting the same number of genes (Fig. 1a). This suggests that both selective constraints and historical contingencies influence the reductive evolution of metabolic networks. Owing to alternative metabolic pathways in the original *E. coli* network, numerous functionally equivalent minimal networks are possible, even under identical selective conditions. For the same reason, only 55% of the reactions are recoverable by single-gene deletion studies (Fig. 1b). The number of genes in the minimal networks is also variable (Fig. 1b), suggesting that there are differences in the number of enzymatic steps between alternative pathways. Deletions at the early stages of genome reduction may affect large genomic regions rather than single genes⁹. However, additional simulations showed that, although allowing such block deletions reduces the number of independent gene-loss events, it has no effect on the size and average similarity of the networks evolved (Supplementary Methods and Supplementary Table 4).

To compare our predictions against real evolutionary outcomes, we divided the *E. coli* enzymes into two mutually exclusive groups: enzymes ubiquitously present in the simulated minimal reaction sets (group A), and enzymes absent in some or all of the simulated sets (group B). If our analysis can approximate reductive evolution in other bacteria, we expect systematic differences in the relative frequencies of these enzymes between species with different lifestyles. As expected, the fraction of enzymes with ubiquitous presence in the simulated minimal reaction sets (group A) is especially high in

¹European Molecular Biology Laboratory, Meyerhofstrasse 1, D-69012 Heidelberg, Germany. ²Department of Zoology, University of Oxford, Oxford OX1 3PS, UK. ³Faculty of Life Sciences, The University of Manchester, Michael Smith Building, Oxford Road, Manchester M13 9PT, UK. ⁴Department of Biology & Biochemistry, University of Bath, Claverton Down, Bath BA2 7AY, UK. ⁵Department of Medical Chemistry, Semmelweis University, PO Box 260, H-1444 Budapest, Hungary.

*These authors contributed equally to this work.

intracellular parasites and endosymbionts as compared with free-living microbes (Fig. 1c).

To investigate further how accurately the model describes reductive evolution in nature, we focused our simulations on three fully sequenced genomes of *B. aphidicola* strains^{10–12} and *W. glossinidia*¹³. These are close relatives of *E. coli* with an evolved intracellular

endosymbiotic lifestyle. Gene acquisition must have been a negligible factor in the evolution of these lineages (Supplementary Methods), providing a unique opportunity to study reductive evolution. Setting boundary conditions that mimic the relevant nutrient conditions and selective forces (Supplementary Tables 2 and 3), we performed simulations as described above.

Detailed physiological studies have shown that *Buchnera* supply their aphid hosts with riboflavin¹⁴ and essential amino acids¹⁵ that are lacking in their hosts' diets. To quantify the agreement between our predictions and the observed reductive evolution in *Buchnera*, while considering gene-content variation in simulated minimal genomes, we used a combined measure of sensitivity and specificity¹⁶. For each possible cutoff (that is, the minimal fraction of simulated genomes in which a gene must be present to predict its presence in *Buchnera*), Fig. 2a shows the fraction of true-positive predictions (sensitivity) plotted against the fraction of false-positive predictions (1 – specificity). The area under the resulting curve gives a cutoff-independent measure of predictive accuracy¹⁶. For each of the *Buchnera* strains, the accuracy of the model is ~80% as compared with the 50% expected by chance (Fig. 2a). The above results remain valid when genes putatively transferred horizontally into *E. coli* since its split

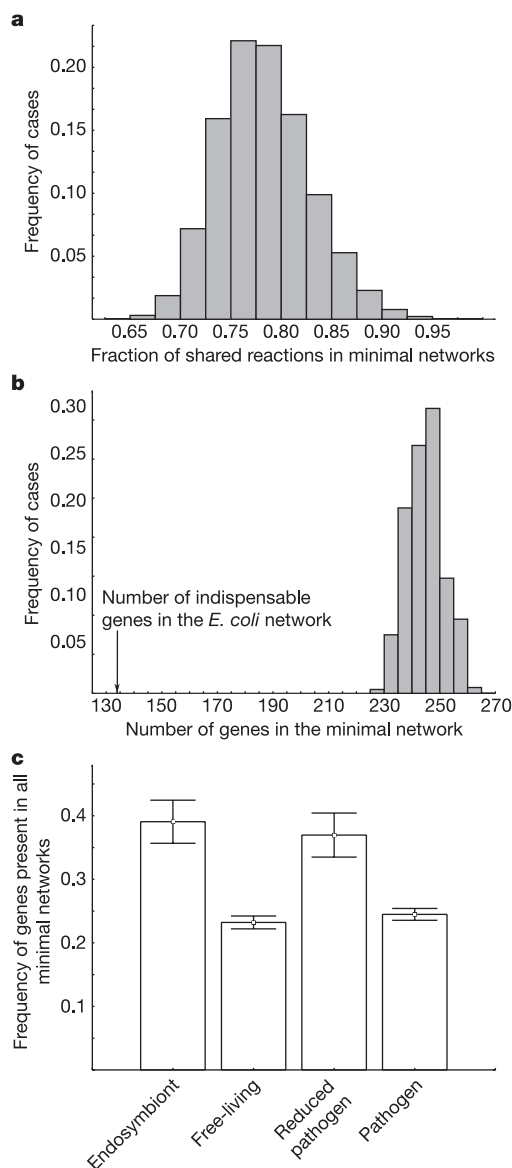


Figure 1 | General properties of evolved minimal networks. **a**, Distribution of the fraction of shared metabolic reactions between all possible pairs among 500 simulated minimal networks. Only reactions with annotated enzyme-encoding genes are shown. The resulting networks share $77 \pm 4.4\%$ (mean $\pm$ s.d.) of their reactions. The 500 networks were generated with random reaction content and the same distribution reaction numbers as the simulants. The average similarity across networks is $25 \pm 2.7\%$. **b**, Distribution of the number of contributing genes in simulated minimal networks. Minimal reaction networks contain, on average, 245 ± 6.48 reactions (mean $\pm$ s.d.); however, only 134 of these genes (~55%) have a predicted fitness effect in the full original *E. coli* network (arrow). **c**, Distribution of genes consistently present in minimal networks in organisms with different lifestyles (Supplementary Table 11). Putative orthologues of *E. coli* enzymes were identified in 140 bacterial species. Shown is the fraction of these that are retained in all simulated minimal networks, summarized across species for each of four different lifestyles (values are the mean $\pm$ 2 s.e.m.). Analysis of variance: $n = 140$, $F = 62.9$, d.f. = 3, $P < 10^{-6}$.

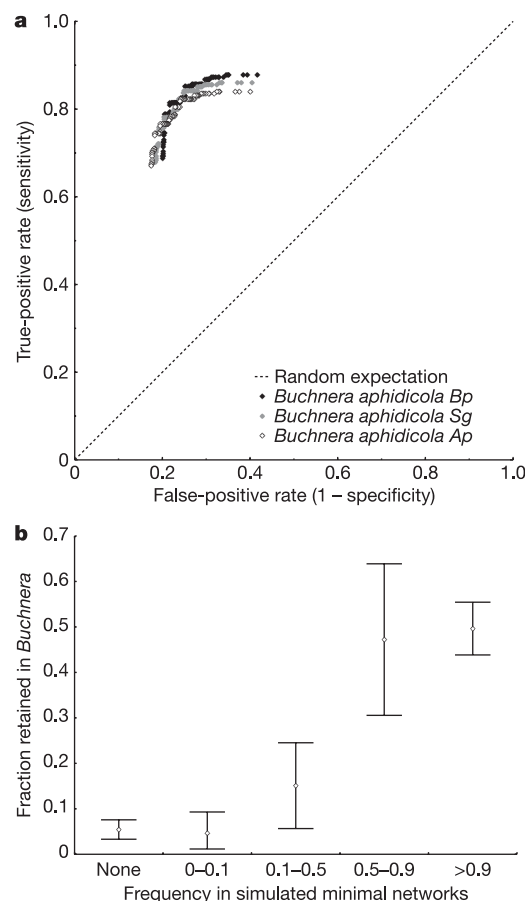


Figure 2 | Comparison of reaction content of simulated and *Buchnera* metabolic networks. **a**, Predictive accuracy for all possible cutoffs (receiver operating characteristic curve)¹⁶. *Bp*: *B. aphidicola*, endosymbiont of *Baizongia pistaciae*; *Sg*: *B. aphidicola*, endosymbiont of *Schizaphis graminum*; *Ap*: *B. aphidicola*, endosymbiont of *Acyrtosiphon pisum*. Overall accuracy (area under curve): *Bp* = 0.802, *Ap* = 0.794, *Sg* = 0.800. All results are highly significant, $P < 10^{-25}$ (see Supplementary Information). **b**, Presence or absence of reactions in *Buchnera aphidicola* *Bp*, averaged over genes within defined ranges of presence or absence in the simulated minimal reaction sets. Error bars indicate 95% confidence intervals. χ^2 -test: $n = 874$, $\chi^2 = 222.6$, d.f. = 4, $P < 10^{-46}$. For results on *Wigglesworthia glossinidia*, see Supplementary Fig. 2.

from the *Buchnera* lineage are excluded from the analysis (Supplementary Methods and Supplementary Table 5). The model also accurately predicts several non-obvious features of *Buchnera* genomes: for example, the retention of particular reactions involved in oxidative phosphorylation and in pyruvate metabolism (Supplementary Table 6).

Consistent with the notion that genes vary widely in their propensity to be lost during reductive evolution, we find a strong correlation between the frequency of a reaction's presence in the simulated reduced networks and its retention in *Buchnera* (Fig. 2b). Metabolic pathways differ widely in their variability across simulated minimal sets (Supplementary Table 7). For example, it seems that there is only one way of producing some key cellular (biomass) components, including compounds for cell wall synthesis and some essential amino acids. By contrast, reactions involved in pyruvate metabolism, nucleotide salvage pathways or transport processes vary in their retention across simulations. For example, there are two distinct pathways by which *E. coli* can activate acetate to acetyl-coenzyme A (ref. 17). These two pathways have been shown experimentally to compensate for deletions in each other in *E. coli*¹⁷, at least under some nutritional conditions. Consistent with this observation, the simulated minimal reaction sets always contain only one of the two pathways; accordingly, *Buchnera* strains have retained only one of the two pathways (Supplementary Table 8).

The above analysis relied on detailed knowledge of the lifestyle of *Buchnera*. Is it possible to predict gene content of an organism with much less information on lifestyle? *Wigglesworthia*, another endosymbiont and close relative of *E. coli*, is an obvious choice. *Wigglesworthia* provides some cofactors and vitamins for its host, the tsetse fly¹⁸. On the basis of the available physiological information¹⁹, it is possible to model the evolution of the metabolic network of this organism with nearly 76% accuracy for the reaction content (Supplementary Fig. 2 and Table 3). It is likely that the available experiments underestimate the number of cofactors produced by the endosymbiont. We thus elaborated a systematic protocol to find the most likely set of cofactors synthesized by *Wigglesworthia* (Supplementary Methods). Based on the idea of greedy algorithms²⁰, the protocol iteratively adds biosynthetic components that must be produced for the host and calculates the impact on the accuracy of predicting the real reaction content of *Wigglesworthia*. In each round, the cofactor resulting in the best prediction is kept and a new round of simulations is started, adding again each of the remaining compounds one at a time (Supplementary Methods). The method substantially increases model accuracy up to 84% (Supplementary Table 5). It also results in a series of non-trivial predictions on the metabolic capability of *Wigglesworthia*. For example, it suggests that this organism retained the ability to synthesize not only protohaem, but also another related cofactor, haem O (Supplementary Methods).

Under a given selection pressure, simulated minimal reactions sets share 82% (*Wigglesworthia*) and 88% (*Buchnera*) of their reactions, respectively. This value drops to 65% when minimal gene sets across different models are compared. This suggests that variability in gene content among species reflects both variation in selection pressures and chance events in the evolutionary history of the endosymbionts (Supplementary Table 9).

Each loss of a reaction reduces the space available for further reductive evolution. This is most obvious for physiologically fully coupled reactions (such as those in linear pathways), which can only fulfil their metabolic function together²¹. As predicted, members of pairs are either lost or retained together in the investigated endosymbionts in 74–84% of cases, whereas only ~50–55% would be expected by chance (Supplementary Table 10).

Deviations between the model predictions and gene content of endosymbionts might be due to incomplete biochemical knowledge or inaccuracies in modelling the types and relative amounts of nutrient conditions and biosynthetic components required by the endosymbiont or the host cell. Finally, hosts and endosymbionts

interact in ways that are not completely understood, and biomass production may be only a rough proxy for endosymbiont fitness. These caveats aside, our approach might be considered a step towards a predictive theory of gene-content evolution. Complementary to traditional approaches, in which lifestyle is inferred from genomic data, it seems possible to take an organism's ecology and to predict which genes it should have by *in silico* network analysis. Moreover, we find that evolutionary paths are contingent on prior gene deletion events, resulting in networks that generally do not represent the most economical solution in terms of the number of genes retained. Thus, history and chance seem to have significant roles not only in adaptive²² but also in reductive evolution of genomes.

These results also have implications for the search for a minimal genome. By using comparative genomics^{23,24} and systematic gene knock-out studies^{25–27}, traditional analyses of minimal gene sets aim to define a repertoire of genes that is necessary and sufficient to support cellular life². The theoretical foundations of the minimal genome concept have remained, however, largely unexplored. We have established that the catalogue of essential genes in free-living species identified by single-gene deletion studies will underestimate the minimal gene set for metabolic system by about 45% (Fig. 1b). Such considerations, and the simulation techniques used to reach these conclusions, should inform attempts by experimentalists to construct minimal genomes by gradual evolution in the laboratory^{28,29}.

METHODS

For full details on orthologue detection and statistical analyses, see Supplementary Methods.

Flux balance analysis of the *E. coli* network. A reconstructed metabolic network (iJR904 GSM/GPR)³ of *E. coli* K12 was used in this study. The model consists of 931 unique biochemical reactions (including transport processes) and 904 genes. The metabolic reconstruction gives accurate information on the stoichiometry and direction of enzymatic reactions, on the presence of isoenzymes, and on enzymatic complexes. Details of flux balance analysis of the *E. coli* metabolic network have been described elsewhere^{4,5}. In brief, it involves two fundamental steps: first, specification of mass balance constraints around intracellular metabolites; and second, maximization of the production of biomass components. The assumption of a steady state of metabolite concentrations specifies a series of linear equations of individual reaction fluxes, which is written in the form $Sv = 0$, where S is the mn stoichiometric matrix (m being the number of metabolites and n being the number of reactions) and v is the vector of individual fluxes through the network. An individual element S_{ij} gives the contribution of the j -th reaction to metabolite i . A biomass reaction describes the relative contribution of metabolites to the cellular biomass. Availability of nutrients and directions of individual reactions were included as boundary conditions (Supplementary Tables 1–3). Using the linear programming package CPLEX 9.0.0, we identified the flux distribution that maximizes the rate of biomass production.

Simulations on reductive evolution. Following previously elaborated protocols⁵, we start by investigating the behaviour of the *E. coli* metabolic network model under a given environmental condition (Supplementary Tables 1–3). Next, we remove a randomly chosen enzyme from the network and calculate the impact of this deletion on the production of biomass components (for a list, see Supplementary Tables 1–3). Enzyme deletions were simulated by constraining the flux of the corresponding reactions to zero and calculating the corresponding knockout flux configuration by established protocols^{4,5}. A gene was classified as having no fitness effect if the biomass production rate of the knockout strain was reduced by less than a given cutoff; different cutoffs led to very similar results (Supplementary Table 5). Deletions of isoenzymes were considered to have no impact on fitness as long as at least one member remained. By contrast, deletion of any of the subunits of a protein complex was considered to result in zero flux through the corresponding reactions. Reactions with no annotated encoding genes were retained throughout the simulations. If the fitness effect of a simulated gene deletion was below the cutoff, the deletion was assumed to be viable and the enzyme was considered to be permanently lost. Otherwise, the gene was restored to the network. The procedure was repeated until no further enzymes could be deleted. This simulation was repeated 500 times; each run provided an independent evolutionary outcome.

The simulations that mimic the evolution of the *Buchnera* metabolic network relied on available biochemical evidence suggesting that glucose and glutamate are the principal carbon sources from which essential amino acids and riboflavin

must be produced for the host (Supplementary Table 2). Besides amino acids, mononucleotides and fatty acids, among others, the biomass components that must be synthesized also include riboflavin. A previous study³⁰ estimated the population size of *Buchnera* as $N_e \approx 10^2$ – 10^3 . Gene deletions are effectively neutral and can thus spread through a population if $|N_e s| < 1$, where s is the selective effect of the gene deletion. Accordingly, the cutoff for the fitness effect of simulated gene deletions was set to 10^{-2} . A less stringent cutoff (0.1) gave very similar results (Supplementary Table 6). For details of *Wigglesworthia* uptake and selective conditions, see Supplementary Table 3.

Received 7 November; accepted 27 December 2005.

1. Tyson, G. W. *et al.* Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**, 37–43 (2004).
2. Koonin, E. V. Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nature Rev. Microbiol.* **1**, 127–136 (2003).
3. Reed, J. L., Vo, T. D., Schilling, C. H. & Palsson, B. O. An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome Biol.* **4**, R54 (2003).
4. Price, N. D., Reed, J. L. & Palsson, B. O. Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nature Rev. Microbiol.* **2**, 886–897 (2004).
5. Edwards, J. S. & Palsson, B. O. The *Escherichia coli* MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. *Proc. Natl Acad. Sci. USA* **97**, 5528–5533 (2000).
6. Gil, R., Latorre, A. & Moya, A. Bacterial endosymbionts of insects: insights from comparative genomics. *Environ. Microbiol.* **6**, 1109–1122 (2004).
7. Klasson, L. & Andersson, S. G. Evolution of minimal-gene-sets in host-dependent bacteria. *Trends Microbiol.* **12**, 37–43 (2004).
8. Burgard, A. P., Vaidyaraman, S. & Maranas, C. D. Minimal reaction sets for *Escherichia coli* metabolism under different growth requirements and uptake environments. *Biotechnol. Prog.* **17**, 791–797 (2001).
9. Moran, N. A. & Mira, A. The process of genome shrinkage in the obligate symbiont *Buchnera aphidicola*. *Genome Biol.* **2**, research0054 (2001).
10. Shigenobu, S., Watanabe, H., Hattori, M., Sakaki, Y. & Ishikawa, H. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. *Nature* **407**, 81–86 (2000).
11. van Ham, R. C. *et al.* Reductive genome evolution in *Buchnera aphidicola*. *Proc. Natl Acad. Sci. USA* **100**, 581–586 (2003).
12. Tamas, I. *et al.* 50 million years of genomic stasis in endosymbiotic bacteria. *Science* **296**, 2376–2379 (2002).
13. Akman, L. *et al.* Genome sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossinidia*. *Nature Genet.* **32**, 402–407 (2002).
14. Nakabachi, A. & Ishikawa, H. Provision of riboflavin to the host aphid, *Acyrtosiphon pisum*, by endosymbiotic bacteria, *Buchnera*. *J. Insect Physiol.* **45**, 1–6 (1999).
15. Baumann, P. *et al.* Genetics, physiology, and evolutionary relationships of the genus *Buchnera*—intracellular symbionts of aphids. *Ann. Rev. Microbiol.* **49**, 55–94 (1995).
16. Hanley, J. A. & McNeil, B. J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **143**, 29–36 (1982).
17. Kumari, S., Tishel, R., Eisenbach, M. & Wolfe, A. J. Cloning, characterization, and functional expression of *acs*, the gene which encodes acetyl coenzyme A synthetase in *Escherichia coli*. *J. Bacteriol.* **177**, 2878–2886 (1995).
18. Zientz, E., Dandekar, T. & Gross, R. Metabolic interdependence of obligate intracellular bacteria and their insect hosts. *Microbiol. Mol. Biol. Rev.* **68**, 745–770 (2004).
19. Nogge, G. Significance of symbionts for the maintenance of an optimal nutritional state for successful reproduction in haematophagous arthropods. *Parasitology* **82**, 101–104 (1981).
20. Corman, T. H., Leiserson, C. E., Rivest, R. L. & Stein, C. *Introduction to Algorithms* (MIT Press, Cambridge, MA, 2001).
21. Burgard, A. P., Nikolaev, E. V., Schilling, C. H. & Maranas, C. D. Flux coupling analysis of genome-scale metabolic network reconstructions. *Genome Res.* **14**, 301–312 (2004).
22. Travisano, M., Mongold, J. A., Bennett, A. F. & Lenski, R. E. Experimental tests of the roles of adaptation, chance, and history in evolution. *Science* **267**, 87–90 (1995).
23. Mushegian, A. R. & Koonin, E. V. A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc. Natl Acad. Sci. USA* **93**, 10268–10273 (1996).
24. Gil, R., Silva, F. J., Pereto, J. & Moya, A. Determination of the core of a minimal bacterial gene set. *Microbiol. Mol. Biol. Rev.* **68**, 518–537 (2004).
25. Westers, H. *et al.* Genome engineering reveals large dispensable regions in *Bacillus subtilis*. *Mol. Biol. Evol.* **20**, 2076–2090 (2003).
26. Kolisnychenko, V. *et al.* Engineering a reduced *Escherichia coli* genome. *Genome Res.* **12**, 640–647 (2002).
27. Hutchison, C. A. *et al.* Global transposon mutagenesis and a minimal *Mycoplasma genome*. *Science* **286**, 2165–2169 (1999).
28. Nilsson, A. I. *et al.* Bacterial genome size reduction by experimental evolution. *Proc. Natl Acad. Sci. USA* **102**, 12112–12116 (2005).
29. Oliver, S. G. From DNA sequence to biological function. *Nature* **379**, 597–600 (1996).
30. Mira, A. & Moran, N. A. Estimating population size and transmission bottlenecks in maternally transmitted endosymbiotic bacteria. *Microb. Ecol.* **44**, 137–143 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. von Mering for providing early access to the updated STRING database. C.P., B.P. and P.C. are supported by the Hungarian Scientific Research Fund (OTKA). C.P. is also supported by an EMBO Long-term Fellowship. B.P. is a Fellow of the Human Frontier Science Program. M.J.L. acknowledges financial support by the Deutsche Forschungsgemeinschaft. Work on systems biology in S.G.O.'s laboratory is supported by the Biotechnology and Biological Sciences Research Council.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to L.D.H. (l.d.hurst@bath.ac.uk).

Analysis of the DNA sequence and duplication history of human chromosome 15

Michael C. Zody¹, Manuel Garber¹, Ted Sharpe¹, Sarah K. Young¹, Lee Rowen², Keith O'Neill¹, Charles A. Whittaker^{1†}, Michael Kamal¹, Jean L. Chang¹, Christina A. Cuomo¹, Ken Dewar^{1†}, Michael G. FitzGerald¹, Chinnappa D. Kodira¹, Anup Madan^{2†}, Shizhen Qin², Xiaoping Yang¹, Nissa Abbasi^{2†}, Amr Abouelleil¹, Harindra M. Arachchi¹, Lida Baradarani^{2†}, Brian Birditt^{2†}, Scott Bloom², Toby Bloom¹, Mark L. Borowsky¹, Jeremy Burke², Jonathan Butler¹, April Cook¹, Kurt DeArellano¹, David DeCaprio¹, Lester Dorris III¹, Monica Dors², Evan E. Eichler³, Reinhard Engels¹, Jessica Fahey^{2†}, Peter Fleetwood^{2†}, Cynthia Friedman⁴, Gary Gearin¹, Jennifer L. Hall¹, Grace Hensley^{2†}, Ericka Johnson^{2†}, Charlien Jones¹, Asha Kamat¹, Amardeep Kaur², Devin P. Locke³, Anuradha Madan^{2†}, Glen Munson¹, David B. Jaffe¹, Annie Lui¹, Pendexter Macdonald¹, Evan Mauceli¹, Jerome W. Naylor¹, Ryan Nesbitt², Robert Nicol¹, Sinéad B. O'Leary¹, Amber Ratcliffe^{2†}, Steven Rounsley¹, Xinwei She³, Katherine M. B. Sneddon⁵, Sandra Stewart², Carrie Sougnez¹, Sabrina M. Stone¹, Kerri Topham¹, Dascena Vincent^{2†}, Shunguang Wang¹, Andrew R. Zimmer¹, Bruce W. Birren¹, Leroy Hood², Eric S. Lander¹ & Chad Nusbaum¹

Here we present a finished sequence of human chromosome 15, together with a high-quality gene catalogue. As chromosome 15 is one of seven human chromosomes with a high rate of segmental duplication¹, we have carried out a detailed analysis of the duplication structure of the chromosome. Segmental duplications in chromosome 15 are largely clustered in two regions, on proximal and distal 15q; the proximal region is notable because recombination among the segmental duplications can result in deletions causing Prader-Willi and Angelman syndromes^{2,3}. Sequence analysis shows that the proximal and distal regions of 15q share extensive ancient similarity⁴. Using a simple approach, we have been able to reconstruct many of the events by which the current duplication structure arose. We find that most of the intrachromosomal duplications seem to share a common ancestry. Finally, we demonstrate that some remaining gaps in the genome sequence are probably due to structural polymorphisms between haplotypes; this may explain a significant fraction of the gaps remaining in the human genome.

The present work describes the completion of a physical map, high-quality finished sequence, and gene catalogue for the euchromatic q arm of human chromosome 15, representing 2.9% of the human genome. The finished sequence contains 81,871,010 bases and is interrupted by nine euchromatic gaps and one gap containing the heterochromatic p arm and centromere regions (Fig. 1). The total size of the euchromatic gaps is estimated at 544 kilobases (kb) (Methods and Supplementary Table S1). These gaps remain despite the screening of genomic libraries containing a combined ~53-fold

physical coverage, and are refractory to current cloning and mapping technology; six are within or adjacent to large duplicated regions. Of the finished sequence, 74% was generated by the Broad Institute of MIT and Harvard (formerly the Whitehead Institute/MIT Center for Genome Research (WICGR)), 25% by the Multimegabase Sequencing Center (initially at the University of Washington, currently at the Institute for Systems Biology), and the remaining ~1% by three other groups (Supplementary Table S2). The analyses here are referenced to NCBI Build 35; however, we have slightly improved this sequence (including closing one of the euchromatic gaps), and provide the updated clone path in Supplementary Table S3. Details of construction of the clone map and sequencing are described in the Supplementary Information. The short arm of chromosome 15, as in other acrocentric human chromosomes (chromosomes 13, 14, 21 and 22), is heterochromatic and was not sequenced as part of the Human Genome Project; it is estimated at 17 Mb (ref. 5) and contains arrays of ribosomal RNA genes, satellite sequences and other repeated sequences⁶.

We assessed the local accuracy of the clone path by aligning paired-end sequences from a human fosmid library (designated WIBR2, representing 10 × physical coverage) to the finished sequence^{7,8}. This analysis revealed no aberrant clones. In addition, an independent quality assessment exercise commissioned by the National Human Genome Research Institute⁹ estimated the accuracy of the finished sequence to be better than one error in 100,000 bases (J. Schmutz, personal communication).

Several analyses suggest that nearly the entire euchromatic region

¹Broad Institute of MIT and Harvard, 320 Charles Street, Cambridge, Massachusetts 02141, USA. ²Institute for Systems Biology, 1441 North 34th Street, Seattle, Washington 98103, USA. ³Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA. ⁴Division of Human Biology, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, Washington 98109, USA. ⁵HUGO Gene Nomenclature Committee (HGNC), The Galton Laboratory, Department of Biology, University College London, Wolfson House, 4 Stephenson Way, London NW1 2HE, UK. [†]Present addresses: MIT Center for Cancer Research, 77 Massachusetts Avenue E18-570, Cambridge, Massachusetts 02139, USA (C.A.W.); McGill University and Genome Quebec Innovation Centre, Montreal, Quebec H3A 1A4, Canada (K.D.); Neurogenomics Research Lab, 200 B EMBR, University of Iowa, Iowa City, Iowa 52242, USA (Anup Madan); Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, Washington 98109, USA (N.A.); Blue Heron Technologies, Bothell, Washington 98021, USA (L.B.); Department of Microbiology, Box 358070, University of Washington, Seattle, Washington 98195, USA (B.B.); Seattle University School of Nursing, Seattle, Washington 98122, USA (J.F.); Corbis Corporation, Seattle, Washington 98104, USA (P.F.); Geospiza, 100 West Harrison North Tower, Suite 330, Seattle, Washington 98119, USA (G.H.); Division of Medical Genetics, Box 357720, University of Washington, Seattle, Washington 98195, USA (E.J.); 3095 Medical Laboratories, Department of Neurosurgery, University of Iowa, Iowa City, Iowa 52242, USA (Anuradha Madan); Nanostring Technologies, 201 Elliott Avenue West, Suite 300, Seattle, Washington 98119, USA (A.R.); Genelex Corporation, 3000 1st Avenue, Suite 1, Seattle, Washington 98121, USA (D.V.).

of chromosome 15 is present and accurately represented in the finished sequence. All genes in the RefSeq¹⁰ database (596 loci, 742 transcripts) previously mapped to chromosome 15 are present and complete in the finished sequence. Furthermore, the finished sequence shows excellent alignment to genetic and radiation hybrid maps (Supplementary Fig. S1). The genetic map¹¹ shows perfect alignment, with no discrepancies among 125 sequence-based genetic markers (Supplementary Table S4). The radiation hybrid map¹² contains only local discrepancies, owing to its lower resolution (Supplementary Table S5). A large gap in the radiation hybrid coordinates (254–280 cR) at ~74 Mb in the physical map, near a

region where chromosome breakage has been observed independently in multiple mammalian lineages (see below), is probably the result of non-random breakage in the generation of the radiation hybrid panel.

We produced a manually curated⁸ catalogue of genes, containing 695 gene loci (including all genes in RefSeq) and 250 pseudogene loci on chromosome 15. Table 1 classifies the genes according to standardized categories. The 3% of genes in the 'novel' and 'putative' categories were annotated based only on spliced expressed-sequence-tag (EST) evidence; some of these may prove to be pseudogenes. The full-length transcripts of known genes have an average length of 3,267 bp, with an average of 11.6 exons. Internal exon lengths average 156 bp. Gene loci have an average of 4.6 distinct transcripts, with 66% having at least two transcripts. These gene statistics are similar to recent reports^{8,13–16}. Examples of genes that represent extremes of these distributions are described in the Supplementary Information. Most (74%) of the 250 pseudogenes are processed. In addition, we identified 9 transfer RNA genes (Supplementary Table S6) and found six known microRNAs mapping to chromosome 15 (Supplementary Table S7).

In most aspects of its landscape, chromosome 15 is close to genome-wide averages⁷. The overall gene density is 8.6 genes per Mb. There are 18 gene deserts (defined as 500 kb without an identified coding gene, Supplementary Table S8) comprising 14.9 Mb (~18.3% of the chromosome). The overall G+C content is 42.2%, but varies substantially across the chromosome (Fig. 1b). Transposable element fossils cover 38.3%. Chromosome 15 is also typical in its content of non-coding sequence conservation (see Supplementary Information).

Chromosome 15 is, however, one of seven autosomes that are significantly enriched in segmental duplications (defined as regions >1 kb that are not high-copy repeats and have >90% identity to another region in the genome¹⁷), with 8.8% of its euchromatin composed of such sequence (Supplementary Fig. S2). As with other heavily duplicated chromosomes, chromosome 15 has a large fraction of intrachromosomal duplication: 50% is strictly intrachromosomal, 30% is both intra- and interchromosomal, and 20% is solely interchromosomal (largely in the proximal 1.5 Mb). The proportion of purely interchromosomal duplication might be even lower, as some undetected tandem duplication may exist near the centromere (see below). Recombination among segmental duplications within the region 15q11–q13 gives rise to deletions that are known to cause Prader-Willi and Angelman syndromes^{2,3} (Supplementary Information).

We sought to investigate the duplication landscape of chromosome 15 by studying the relationships among the duplicated segments. Previous work has shown that a sequence within the Prader-Willi/Angelman syndrome region, termed LCR15 (ref. 4), is also duplicated on distal 15q (Supplementary Fig. S2). By extending our analysis to detect more ancient relationships (sequence identity less than 90%), we found much more extensive similarity among the duplicated sequences in both proximal and distal 15q (Fig. 1a). We clustered together segmental duplications containing related sequence (Methods) and found that most fell into a single large cluster, which we refer to as 'class 1'. The class includes 67% of all bases in segmental duplications and 91% of all pairwise duplication events (as some bases reside within multiple independent events) (Supplementary Table S9).

Although the segmental duplications are related to one another in a complex fashion, we sought to identify a 'core element' that was present in many of the class 1 elements. We took the longest duplicated class 1 region (213 kb starting at 18.89 Mb, within the Prader-Willi/Angelman syndrome region) and aligned all duplicated regions of the chromosome to it, counting the number of different duplication regions that aligned to each base. We selected a core element that includes the highest peak of coverage (Supplementary Fig. S3); the element is 2,920 bp long and lies within the ~15-kb LCR15 element.

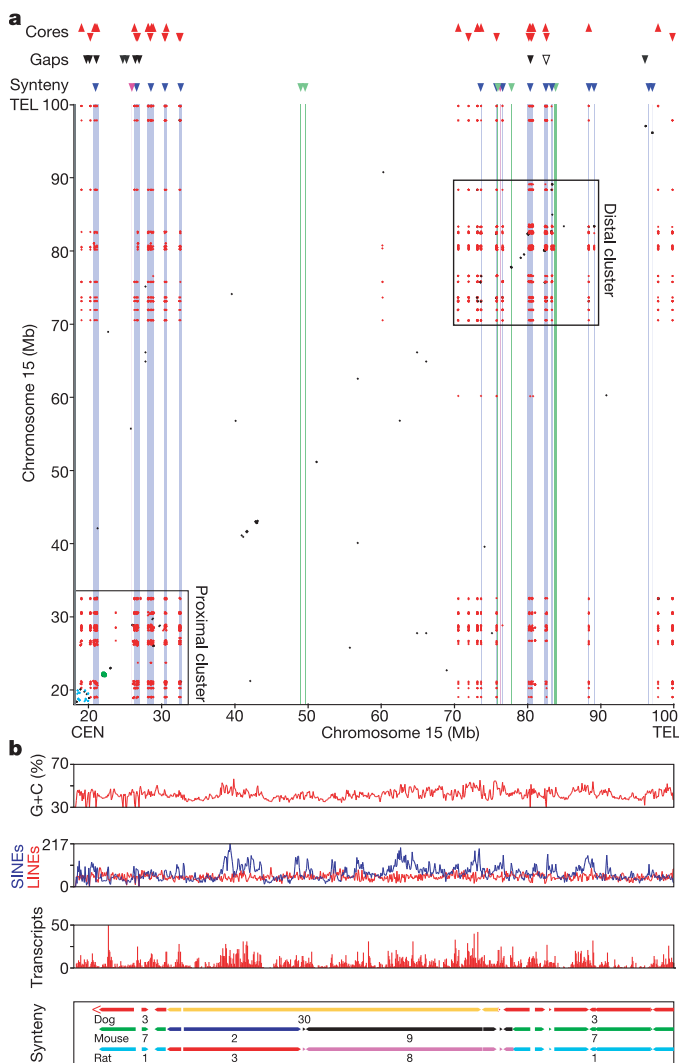


Figure 1 | Overview and duplication content of human chromosome 15.

a, Dot-plot of duplicons on human chromosome 15, showing association with species-specific breaks in conserved synteny. Class 1 duplications are shown in red; other coloured dots indicate alignments between minor duplication classes. Vertical bands topped by arrows represent breaks in synteny (human-specific in purple, rodent-specific in green, dog-specific in pink). Black arrows at the top denote gaps in the human sequence (open arrow indicates a gap that was closed after Build 35 was made). Red triangles at the top show the locations and strand of class 1 core elements. The 15q telomere (TEL) and the centromere (CEN) are indicated. **b**, The following features are represented in discrete windows of 100 kb (top to bottom): G+C content on a scale from 30–70%; densities of LINES (red) and SINES (blue) (long and short interspersed elements, respectively); and transcripts as counts of elements. The bottom panel shows blocks of conserved synteny (100-kb resolution) with dog, mouse and rat. Chromosomes are numbered, and are coloured arbitrarily for ease of distinction.

Table 1 | Chromosome 15 gene content

Category	Gene number	Gene percentage	Gene length (bp)*	Number of alternative transcripts	Transcript length (bp)†	Number of exons per transcript‡	Internal exon length (bp)§	Intron length (bp)	CpG-5' association¶
Known genes	532	76	66,994	4.6	3,267	11.6	156 (n = 6,471)	6,157 (n = 8,277)	76
Novel CDS	73	11	40,090	2.1	1,185	5.2	154 (n = 278)	8,108 (n = 384)	35
Novel transcripts	68	10	29,855	1.8	867	3.5	146 (n = 182)	8,851 (n = 351)	46
Putative genes	15	2	10,074	1.5	1,070	2.9	109 (n = 10)	6,700 (n = 35)	47
Gene fragments	7	1	1,563	1.0	425	2.3			
Total	695		57,963						
Pseudogenes	250	26	3,297	1.0	1,091	2.2	195 (n = 234)	1,878 (n = 294)	27

Categorization is according to Hawk2 standards (<http://www.sanger.ac.uk/Info/workshops/hawk2>; see Supplementary Information). CDS, coding sequences.
* Average chromosomal distance from the beginning of the 5'-most exon to the end of the 3'-most exon in all transcripts in a gene.
† Average length summed across the footprint of all exons in all transcripts in a gene (total exon space per gene).
‡ Average number of exons in transcripts. Exons common to different transcripts were counted once per transcript.
§ Average length of exons using the footprint of all non-terminal exons of all transcripts in a gene. Unique overlapping exons or contained exons are counted separately, making this an average length of unique exons in a gene. (Sample size given in parentheses.)
|| Average length of unique introns in a gene. In the case of exon skipping, both the shorter and longer versions of the overlapping introns were counted towards the average. (Sample size given in parentheses.)
¶ Percentage of genes with a transcript having a CpG island (as assessed by FirstEF) within -2 kb and +1 kb of the transcription start.

The human genome contains 41 nearly full-length copies of the core element: there are 37 on chromosome 15, two on the Y chromosome, and one each on chromosomes 2 and 10. To understand the origins of the element, we compared the core element to the dog¹⁸ and mouse¹⁹ genomes. The dog and mouse genomes each contain a single copy of the element, which is orthologous to the copy on human chromosome 2. The similarity among the sequences is shown in a phylogenetic tree (Fig. 2, see Methods). The copy on chromosome 2 is at the root of the human duplications, closest to mouse and dog, as would be expected from conserved synteny. The duplications on chromosome 15 fall into two distinct and well-separated branches: a proximal branch containing all the elements in the Prader-Willi/Angelman syndrome region (chromosome position 18–32 Mb), and a distal branch containing all the elements from 73 to 88 Mb, with a tight clustering of elements around 80–83 Mb. A further two repeats in the subtelomeric region (98–100 Mb) are closely related to the proximal branch. Pairwise divergence between elements in the two branches is ~11%, indicating that they share an ancient origin followed by local duplications, but with no recent interaction between branches.

From the tree, it is possible to reconstruct the likely history of the core element. The sequence on chromosome 2 lies in the 3' untranslated region (UTR) of a splice variant of the gene *intersectin 2* (*ITSN2*). This sequence seems to have moved by retroposition to chromosome 10 (at 30.68 Mb), inserting immediately downstream of the 5' coding sequence of an interchromosomally duplicated copy of *GOLGA2* (the origin of which is on chromosome 9). A combined unit (15 kb, consisting of *GOLGA2* and the *ITSN2* UTR) then was copied to chromosome 15, where it has duplicated extensively. Finally, two copies exist on the arms of a large palindrome on the Y chromosome, and seem to have moved to the Y chromosome by segmental duplication of ~40 kb of chromosome 15 (at 82.7 Mb).

We next sought to understand why the large regions of segmental duplication in proximal 15q (denoted 'A') and distal 15q (denoted 'C') are separated by a large stretch that contains almost no duplicated sequence (denoted 'B'). Analysis of conserved synteny with other species allows a reconstruction of the history of chromosome 15 (Fig. 3). Briefly, the three segments were adjacent in the boreoeutherian ancestor (the common ancestor of Euarchontoglires and Laurasiatheria), but were found in the order A–C–B. In the primate lineage, the chromosome apparently underwent a single large inversion that separated segments A and C. (Details of the reconstruction and comparison to recent reports^{20,21} can be found in the Supplementary Information and Supplementary Fig. S4.) This suggests that the core element was transferred to chromosome 15 before the divergence of apes and Old World monkeys, and expanded locally (in the originally contiguous A–C region). The inversion subsequently separated regions A and C, and the element continued to expand separately in each region.

To test this hypothesis, we examined the current draft assembly of the rhesus macaque genome (rheMac1; R. Gibbs, personal communication). We found at least 12 nearly full-length copies of the core element that we added to the evolutionary tree (Supplementary Fig. S5). We also found unique orthologues of the copies on human chromosomes 2 and 10. The remaining macaque elements were split between the proximal and distal clusters, confirming that the element

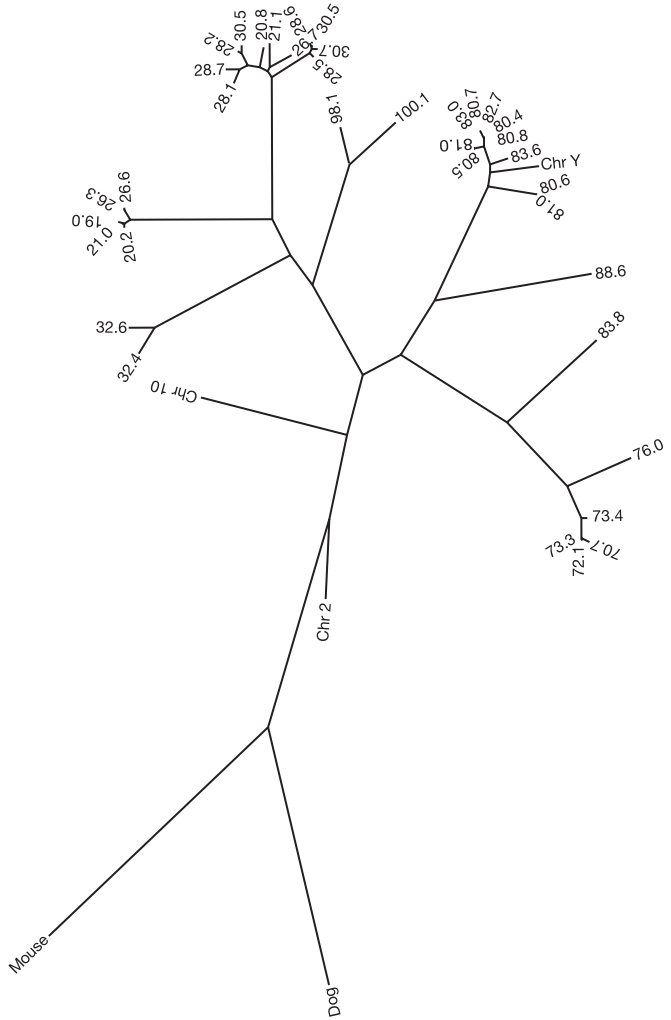


Figure 2 | Phylogenetic tree of the 41 human copies and the unique dog and mouse copies of the conserved core element. Chromosome 15 copies are distinguished by their physical position (in Mb). Chr, chromosome.

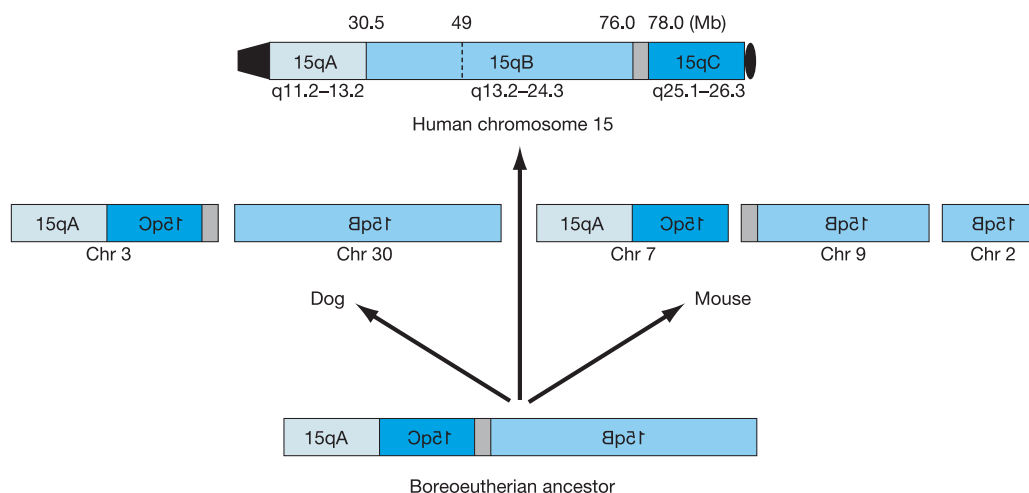


Figure 3 | History of the structural organization of human chromosome 15. For simplicity, we have depicted the q arm as three large segments—A, B and C—that have travelled together. Internal rearrangements exist within these segments, but do not cross between them in mammals. Rat (not

shown) is identical to mouse except for the chromosome numbers. The breakpoints between distal B and proximal C differ by 2 Mb in rodent and dog (grey box). Segments labelled with backwards text are inverted with respect to the modern human chromosome. Chr, chromosome.

had already appeared and begun to duplicate on chromosome 15 before the divergence of Old World monkeys and apes. The human and macaque elements are grouped into separate clusters in both the proximal and distal branches, indicating that local duplication has continued to occur in both the human and macaque lineages.

The analysis of conserved synteny also reveals that the segmental duplications are closely associated with chromosomal rearrangements. Chromosome 15 has 15 human-specific breakpoints of conserved synteny, all of which are inversions. Of these, 13 occur in regions containing class 1 duplications. This suggests that the segmental duplications may have mediated the inversions and that these inversions may have helped to disperse the elements.

The class 1 core element serves as a useful marker for tracing chromosomal history. However, the ubiquity of the core element raises the possibility that it had a causal role in the process of segmental duplication on chromosome 15. The element is derived from a UTR on chromosome 2, of which at least 500 bases are highly conserved across mammals and thus are presumably functional. Moreover, many of the copies on chromosome 15 are transcribed: 13 known genes on chromosome 15 (all golgins or golgin-like proteins) contain this duplicated UTR, and another 16 transcripts stop just short of it (Supplementary Table S10). It will be interesting to investigate whether functional properties of the fusion element on chromosome 15 promote local duplication, and to explore whether this had significant implications for primate evolution.

Finally, we note that the segmental duplications represent the main challenge in closing the remaining gaps in the sequence of chromosome 15. Build 35 contains ten gaps, seven of which lie within or immediately adjacent to class 1 duplications (Fig. 1). In some cases, the duplicated sequences flanking the gaps are so similar (>99.7% identity) that they may represent allelic variants. Moreover, six of the seven duplication-associated gaps are adjacent to or within reported sites of copy-number polymorphism^{22,23} (Supplementary Table S1). We have recently been able to close one gap (at 82.7 Mb) (decreasing the number of gaps to nine) by finding previously missed overlap between two flanking clones; another clone spanning this gap carries an alternative haplotype with an additional 100 kb, including an 80-kb near-perfect duplication. Examination of three of the other gaps suggests that they might also be due to structural variation, although more work will be required to confirm this.

The finished sequence of chromosome 15 offers a window into the natural history of segmental duplications and the structural

history of chromosomes. Notably, most of the intrachromosomal duplication involves a single class of duplicons. On the basis of these results, we suggest an important role for such duplicons in structural evolution and gene diversification.

METHODS

Production of gene catalogue and annotation. The gene catalogue was produced as described previously⁸. Gene symbols were assigned by the HUGO Gene Nomenclature Committee for biologically characterized loci. A complete list of gene symbols from this paper can be found in Supplementary Table S11. Annotation was performed as described previously⁸. Our annotations are available from the Vertebrate Genome Annotation database (VEGA, http://vega.sanger.ac.uk/Homo_sapiens).

Segmental duplications. Segmental duplications were defined as pairs of regions of 90% or greater identity (excluding repeat-masked bases) that extend for 1 kb or more. The map of segmental duplications was prepared using a method adapted from ref. 17, by concatenating all-against-all MegaBlast²⁴ alignments. A genome database was built using hard-masked sequence. This same hard-masked sequence was presented to MegaBlast as a probe, chromosome by chromosome. All alignments of 80% or better identity with expectation <10⁻⁴ were kept. Alignments were then concatenated if they were contiguous except for masked repeats. Unmasked gaps could be crossed but were penalized to prevent over-merging by being treated as bases of 50% identity. Final segments meeting the 1-kb length and 90% identity criteria were retained.

Duplication class clustering. Pairwise intrachromosomal duplications were defined as above. A pairwise duplication A ~ A' was considered to be in the same class as another pairwise duplication B ~ B' if B or B' overlapped A or A' by 150 bp or more. We extended this by transitive closure to build maximally linked sets (that is, if A ~ A' linked to B ~ B' and C ~ C', all were clustered, even if B ~ B' did not overlap C ~ C'). The number of duplications in a class is counted as the number of distinct pairwise alignments X ~ X' that were clustered. The number of bases in a class is counted as the number of distinct bases covered by at least one pairwise duplication in that class.

Construction of core element phylogeny. Full-length or nearly full-length copies of the core element in human were identified by MegaBlast (release 2.2.11). Copies in the mouse and dog genomes were identified by MegaBlast followed by blastn (release 2.2.11) to refine the boundaries and extend the regions. Multiple alignments of the elements were generated with ClustalW (v.1.83). Pairwise and multiple alignment parameters were adjusted by reducing the gap extension penalty to 0.1 and replacing the standard DNA matrix with a custom matrix scoring 10 for any match, -5 for any mismatch, and 0 for any alignment to an unknown base (N). The trees were output in phylip format and all gaps of length >1 converted to single indels by substitution of '?' characters for all but the first '-' in the gap to avoid generating disproportionately long branches for element copies with substantial deletion. Terminal gaps were also treated this way. Trees were built with the dnaps parsimony module of phylip (v.3.65)²⁵. The tree represented is the first of 15 equally likely trees that differ only

in the leaf placement of the seven nearly identical copies of the element at 80 and 82 Mb on chromosome 15.

Received 9 November 2005; accepted 26 January 2006.

1. Bailey, J. A. *et al.* Recent segmental duplications in the human genome. *Science* **297**, 1003–1007 (2002).
2. Emanuel, B. S. & Shaikh, T. H. Segmental duplications: an 'expanding' role in genomic instability and disease. *Nature Rev. Genet.* **2**, 791–800 (2001).
3. Soejima, H. & Wagstaff, J. Imprinting centers, chromatin structure, and disease. *J. Cell. Biochem.* **95**, 226–233 (2005).
4. Pujana, M. A. *et al.* Additional complexity on human chromosome 15q: Identification of a set of newly recognized duplicons (LCR15) on 15q11–q13, 15q24, and 15q26. *Genome Res.* **11**, 98–111 (2001).
5. Morton, N. Parameters of the human genome. *Proc. Natl Acad. Sci. USA* **88**, 7474–7476 (1991).
6. Kehrer-Sawatzki, H. *et al.* Mapping of members of the low-copy-number repetitive DNA sequence family from chAB4 within the p arms of human acrocentric chromosomes: characterization of Robertsonian translocations. *Chromosome Res.* **6**, 429–435 (1998).
7. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931–945 (2004).
8. Nusbaum, C. *et al.* DNA sequence and analysis of human chromosome 18. *Nature* **437**, 551–555 (2005).
9. Schmutz, J. *et al.* Quality assessment of the human genome sequence. *Nature* **429**, 365–368 (2004).
10. Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **33**, D501–D504 (2005).
11. Kong, A. *et al.* A high-resolution recombination map of the human genome. *Nature Genet.* **31**, 241–247 (2002).
12. Schuler, G. D. *et al.* A gene map of the human genome. *Science* **274**, 540–546 (1996).
13. Nusbaum, C. *et al.* DNA sequence and analysis of human chromosome 8. *Nature* **439**, 331–335 (2006).
14. Hillier, L. W. *et al.* Generation and annotation of the DNA sequences of human chromosomes 2 and 4. *Nature* **434**, 724–731 (2005).
15. Grimwood, J. *et al.* The DNA sequence and biology of human chromosome 19. *Nature* **428**, 529–535 (2004).
16. Deloukas, P. *et al.* The DNA sequence and comparative analysis of human chromosome 10. *Nature* **429**, 375–381 (2004).
17. Bailey, J. A. *et al.* Segmental duplications: organization and impact within the current human genome project assembly. *Genome Res.* **11**, 1005–1017 (2001).
18. Lindblad-Toh, K. *et al.* Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* **438**, 803–819 (2005).
19. Mouse Genome Sequencing Consortium. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
20. Murphy, W. J. *et al.* Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps. *Science* **309**, 613–617 (2005).
21. Bourque, G., Zdobnov, E. M., Bork, P., Pevzner, P. A. & Tesler, G. Comparative architectures of mammalian and chicken genomes reveal highly variable rates of genomic rearrangements across different lineages. *Genome Res.* **15**, 98–110 (2005).
22. Sharp, A. J. *et al.* Segmental duplications and copy-number variation in the human genome. *Am. J. Hum. Genet.* **77**, 78–88 (2005).
23. Sebat, J. *et al.* Large-scale copy number polymorphism in the human genome. *Science* **305**, 525–528 (2004).
24. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *NAR* **25**, 3389 (1997).
25. Felsenstein, J. PHYLIP—Phylogeny Inference Package (version 3.2). *Cladistics* **5**, 164–166 (1989).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank L. Gaffney for help with figures and text. We are grateful to T. Furey for help with lists of genetic markers and placement of RefSeqs, and to K. Lindblad-Toh for sharing data from the genome projects of dog and opossum. Fluorescence *in situ* hybridization (FISH) data for opossum were provided by M. Breen. We thank the members of the Baylor College of Medicine Human Genome Sequencing Center, the J. Craig Venter Institute Joint Technology Center, and the Washington University Genome Sequencing Center for generation and early release of the assembly of the rhesus macaque genome. We thank the Sanger Institute for gap sizing by FISH. We also acknowledge the HUGO Gene Nomenclature Committee (S. Povey (chair), E. A. Bruford, V. K. Khodiyar, R. C. Lovering, M. J. Lush, T. P. Sneddon, C. C. Talbot Jr and M. W. Wright) for assigning official gene symbols. We are grateful to all members, present and past, of the Broad (and Whitehead) sequencing platform for their dedication and the consistent high quality of their data.

Author Information Accession numbers for all clones contributing to the finished sequence of human chromosome 15 can be found in Supplementary Table S3. The updated human chromosome 15 sequence can be accessed through GenBank accession number NC_000015. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.C.Z. (mczody@broad.mit.edu) or C.N. (chad@broad.mit.edu).

LETTERS

Intellectual ability and cortical development in children and adolescents

P. Shaw¹, D. Greenstein¹, J. Lerch², L. Clasen¹, R. Lenroot¹, N. Gogtay¹, A. Evans², J. Rapoport¹ & J. Giedd¹

Children who are adept at any one of the three academic 'R's (reading, writing and arithmetic) tend to be good at the others, and grow into adults who are similarly skilled at diverse intellectually demanding activities^{1–3}. Determining the neuroanatomical correlates of this relatively stable individual trait of general intelligence has proved difficult, particularly in the rapidly developing brains of children and adolescents. Here we demonstrate that the trajectory of change in the thickness of the cerebral cortex, rather than cortical thickness itself, is most closely related to level of intelligence. Using a longitudinal design, we find a marked developmental shift from a predominantly negative correlation between intelligence and cortical thickness in early childhood to a positive correlation in late childhood and beyond. Additionally, level of intelligence is associated with the trajectory of cortical development, primarily in frontal regions implicated in the maturation of intelligent activity^{4,5}. More intelligent children demonstrate a particularly plastic cortex, with an initial accelerated and prolonged phase of cortical increase, which yields to equally vigorous cortical thinning by early adolescence. This study indicates that the neuroanatomical expression of intelligence in children is dynamic.

Structural neuroimaging studies generally report a modest correlation ($r = 0.3$) between psychometric measures of intelligence and total brain volume⁶. Links between intelligence and specific regions of the brain may vary according to developmental stage: the anterior cingulate in children⁷, the orbitofrontal and medial prefrontal cortex in adolescents⁸, and the lateral prefrontal cortex in older adults⁹. Most previous studies infer developmental processes from purely cross-sectional data, an endeavour fraught with methodological complications¹⁰. Only one longitudinal study has linked cortical development with cognitive variation, demonstrating greater cortical thinning in the left dorsal frontal and parietal regions among children who gained more in a measure of verbal intelligence⁵. However, this study was limited by its small sample size ($n = 45$), narrow age range (5–11 yr), and consideration of only linear cortical change, whereas brain development generally follows more complex growth patterns^{7,11}.

We characterized brain development from childhood to adulthood in a large group of typically developing subjects ($n = 307$), the majority of who had prospectively acquired repeated neuroanatomic scans (see the Methods). Subjects were stratified on the basis of Wechsler intelligence scales, which give a standardized 'intelligence quotient' (IQ) based on subtests assessing verbal and non-verbal knowledge and reasoning¹². We examined the thickness of the cortex throughout the entire cerebrum, as it is a sensitive index of normal brain development^{5,13}, using a fully automated technique, and have validated these measurements by expert manual determination of cortical thickness and population simulations^{14,15}. We reasoned that the trajectory of cortical development in children stratified on the

basis of IQ would differ primarily in the prefrontal cortex, which has both structural and functional correlations with intelligence. The institutional review board of the National Institutes of Mental Health approved the research protocol, and written informed consent and assent were obtained from parents and children, respectively.

We estimated Pearson's correlations between IQ and cortical thickness for all subjects (each subject contributing one scan), and found modest positive correlations throughout most of the frontal, parietal and occipital cortex, and similarly modest negative correlations in the anterior temporal cortex (Fig. 1 and Supplementary Table 1). Throughout most of the cerebral cortex, the correlations were not significant at an unadjusted $P < 0.05$.

Dividing the sample into different age groups, however, revealed notable age-related changes. A predominantly negative correlation between IQ and cortical thickness in the early childhood group contrasted with later positive correlations, which peaked in late childhood, but were present in an attenuated form in the adolescent and early adult groups. The change in the valence of the correlation between IQ and cortical thickness was significant between the young and late childhood groups throughout the prefrontal cortex, and the

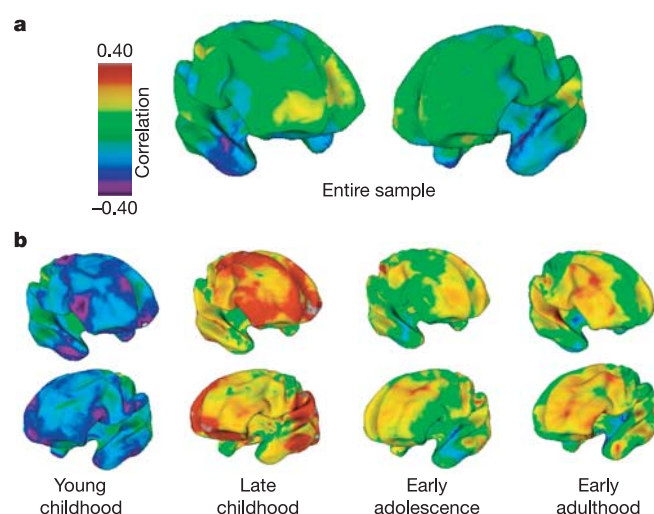


Figure 1 | Correlations between IQ and cortical thickness. **a**, Pearson's correlations for all 307 subjects were generally positive and modest ($P > 0.05$), with r between 0 and 0.10 (green/yellow), except in the anterior temporal cortex (which showed a negative correlation, with r between 0 and -0.1 ; blue/purple). **b**, Correlations in different age groups showed that negative correlations were present in the youngest group, indicating that higher IQ was associated with a thinner cortex particularly in frontal and temporal regions. The relationship reverses in late childhood, with most of the cerebral cortex correlating positively with IQ.

¹Child Psychiatry Branch, National Institute of Mental Health, Bethesda, Maryland 20182, USA. ²Montreal Neurological Institute, McGill University, Montreal, Quebec H3A 2B4, Canada.

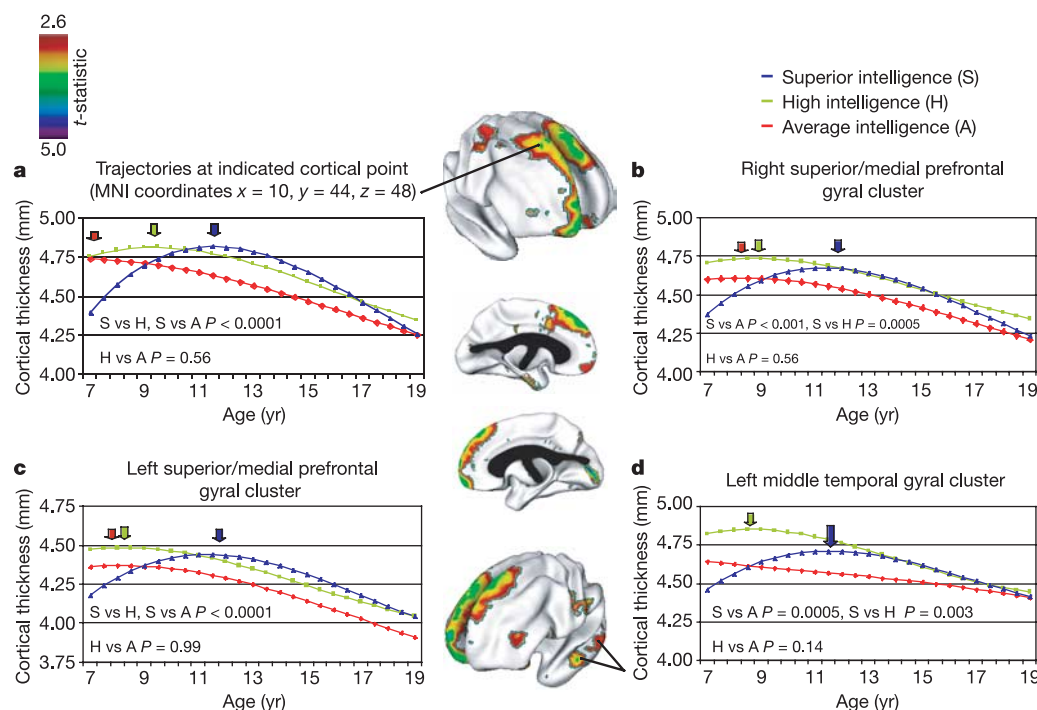


Figure 2 | Trajectories of cortical change. The brain maps (centre panel) show prominent clusters where the superior and average intelligence groups differ significantly in the trajectories of cortical development (t -statistic maps show areas of significant interaction between these IQ groups and the cubic age term). **a**, Graph showing the trajectories at the cortical point of maximum trajectory difference in the right superior frontal gyrus (point

indicated in upper brain map). **b–d**, Graphs showing the trajectories of the mean thickness of all cortical points in the other clusters. The graph in **d** relates to the area indicated in the lower brain map. The age of peak cortical thickness is arrowed and significance values of differences in shapes of trajectories are given on the graphs. MNI, Montreal Neurological Institute.

left superior/middle temporal gyri. These age groups did not differ in gender composition ($\chi^2 = 2.76$; $P = 0.62$) or mean IQ ($F_{3,303} = 1.58$; $P = 0.19$), and there was no significant gender difference in the correlation between cortical thickness and IQ.

We further characterized the development of the relationship between intelligence and cortical morphology using linear mixed-models, which allowed inclusion of all 629 scans. In the determination of cortical thickness, there was a significant interaction between IQ and age terms in the prefrontal cortex, suggesting that the relationship between cortical thickness and IQ varies with age (specifically cubic and quadratic age terms; see the Supplementary Figure).

To explore this interaction, the sample was split into three IQ groups: superior, high and average intelligence. Prominent clusters of cortical points showing differences in cortical development between the intelligence groups lay bilaterally within the superior frontal gyri extending into the medial prefrontal cortex, and to a lesser extent in the middle and orbitofrontal cortices (Fig. 2). In each of these clusters, the trajectories for the local point of maximum trajectory difference and for the entire cluster were similar: the superior intelligence group started from a relatively thinner cortex, but then showed a marked increase in cortical thickness peaking at ~ 11 yr. In contrast, the average intelligence group showed either a steady decline in cortical thickness throughout the age period covered (in orbitofrontal areas), or a short initial increase in cortical thickness which peaked at ~ 7 – 8 yr (in superior frontal gyri). The trajectories of the high intelligence group followed an intermediate pattern, more strongly resembling the pattern of the average intelligence group, with no significant differences between these two groups in the clusters shown in Fig. 2 (all $P > 0.10$).

Different developmental trajectories were also prominent in the posterior left hemisphere between the superior and average intelligence groups, specifically within the left middle prefrontal and inferior temporal gyri, and to a lesser extent the angular gyrus. The

right hemisphere outside the frontal lobes showed trajectories of cortical development that did not differ significantly between groups.

An overall decline in cortical thickness was noted in all groups, present either throughout the age period covered (average intelligence group) or starting by late childhood (high intelligence) or early adolescence (superior intelligence). Velocity curves derived using a first-order differential of the fitted cubic growth curves illustrate that the superior intelligence group had the most rapid rate of cortical thinning, whereas the high and average intelligence groups had similar, but slower, rates (Fig. 3). Thus, the relatively rapid increase in cortical thickness in the superior intelligence group was followed by a more rapid thinning.

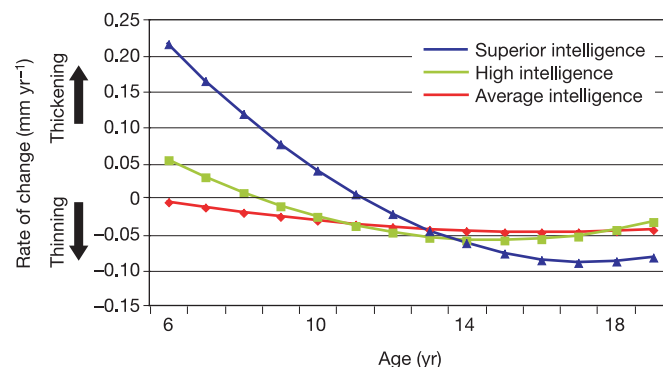


Figure 3 | Rate of change in cortical thickness. The rate of change for the cluster of cortical points in the right superior and medial frontal gyrus, which showed a significant trajectory difference. Positive values indicate increasing cortical thickness, negative values indicate cortical thinning. The point of intersection on the x axis represents the age of maximum cortical thickness (5.6 yr for average, 8.5 yr for high, and 11.2 yr for the superior intelligence group).

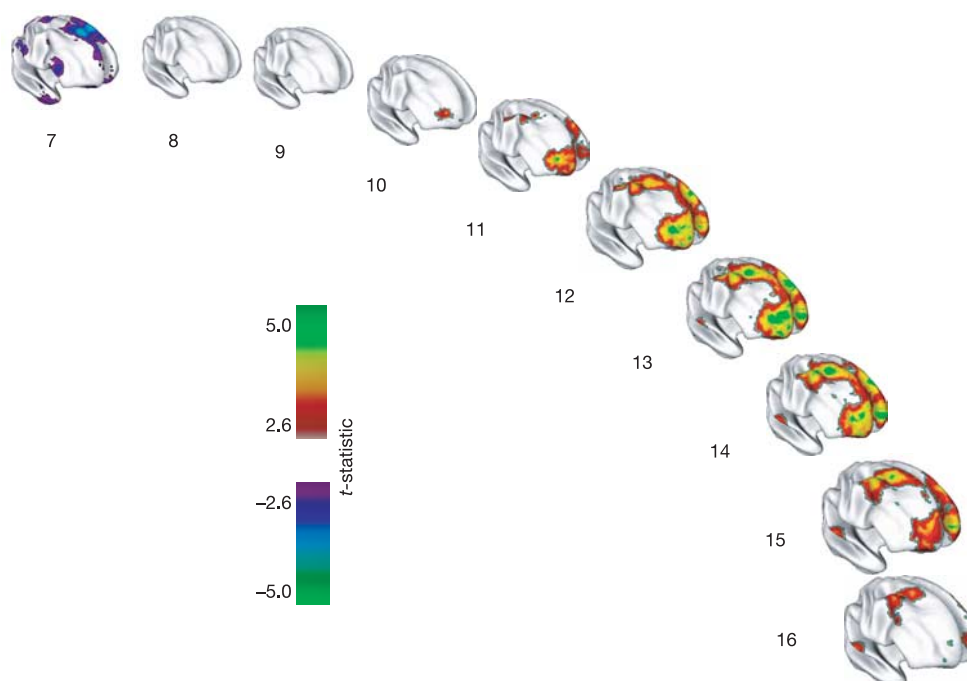


Figure 4 | Developing differences in cortical thickness between the superior and average intelligence groups. Group differences are represented by t -statistics ($t > 2.6$), and show that the superior intelligence group has a thinner superior prefrontal cortex at the earliest age (purple

regions). There is then a rapid increase in cortical thickness (red, green and yellow regions) in the superior intelligence group, peaking at age 13 and waning in late adolescence.

To illustrate the development of differences in cortical thickness between the superior and average intelligence groups, statistical maps representing group differences in the height of the developmental curves at each age were estimated from 7–16 yr (Fig. 4). Initially, the superior intelligence group had a relatively thinner cortex in superior prefrontal gyri, but then showed a rapid increase in cortical thickness. By 11 yr, regions of thicker cortex became apparent in the superior intelligence group—initially in anterior portions of the right superior and middle frontal gyri, spreading to involve more posterior regions of the right prefrontal cortex and the left superior and middle frontal gyri. By late adolescence, the accelerated rate of cortical loss in the most intelligent group leads to decreased regional differences.

The intelligence groups did not differ significantly in handedness or gender composition, but did in socio-economic status ($F_{2,291} = 14.1$; $P < 0.001$), which was correlated with IQ ($r = -0.35$; $P < 0.01$). In the frontal clusters, where trajectories were most closely tied to intelligence, none of these variables contributed significantly to the final polynomial regression model (all P values > 0.1).

Thus, we have demonstrated that level of intelligence is related to the pattern of cortical growth during childhood and adolescence. The differing trajectories of cortical change are most prominent in the prefrontal cortex, congruent with functional magnetic resonance imaging (fMRI) studies showing that activation of the lateral prefrontal cortex is common to a range of intelligence tests, and that the magnitude of frontal cortical activation correlates highly with intelligence^{16,17}.

Our longitudinal structural MRI images provide adequate resolution to describe an *in vivo* change in cortical thickness, but the nature of the underlying cellular events is largely unknown. A determinant of cerebral lamination *in utero* and perinatally is the emergence and resolution of the subplate, which contains neurons, developing cortical afferents and their synapses^{18,19}. Proliferation of myelin into the peripheral cortical neuropil in childhood and adolescence is another possible mechanism influencing cortical thickness^{5,20}. Additionally, the formation and usage-dependent selective elimination of synapses²¹,

which help to create and sculpt neural circuitry including those supporting cognitive abilities²², may contribute to changing cortical dimensions. The prefrontal cortex shows relatively late structural¹¹ and metabolic²³ maturation, and the prolonged phase of prefrontal cortical gain in the most intelligent might afford an even more extended ‘critical’ period for the development of high-level cognitive cortical circuits.

‘Brainy’ children are not cleverer solely by virtue of having more or less grey matter at any one age. Rather, intelligence is related to dynamic properties of cortical maturation.

METHODS

Subjects. Three hundred and seven unrelated children and adolescents with no personal or family history of psychiatric or neurological disorders were recruited (Supplementary Table 2). All subjects had age-appropriate versions of the Wechsler intelligence scales. In 220 subjects, full-scale IQ was estimated from four subtests (vocabulary, similarities, block design and matrix reasoning), and in 87 children two subtests were used (vocabulary and block design). For longitudinal analyses, subjects were divided into three groups on the basis of full-scale IQ with the primary constraint of attaining a roughly equal number of total scans in each group. The groups were: superior intelligence (IQ range 121–149), high intelligence (IQ range 109–120) and average intelligence (IQ range 83–108). All subjects were scanned at least once; 178 participants (58%) had at least two scans; 92 (30%) had three or more scans; the mean interscan interval was ~ 2 yr.

Neuroimaging. T1-weighted magnetic resonance images (1.5 mm axial and 2 mm coronal slices), acquired using three-dimensional spoiled gradient recalled echo in the steady state on a 1.5-T Signa scanner (General Electric), were registered to standardized space²⁴ and corrected for non-uniformity artefacts²⁵. The inner and outer cortical surfaces were extracted from tissue-segmented images using deformable models, and non-linearly aligned towards a standard template surface²⁶. Cortical thickness was measured in native space millimetres using the linked distance between the pial white and grey matter surfaces at 40,962 vertices throughout the cerebral cortex²⁷ (see Supplementary Methods). In order to improve the ability to detect population changes, each cortical thickness map was blurred using a 30-mm surface-based blurring kernel, which respects anatomical boundaries and was chosen to maximize statistical power while minimizing false positives¹⁵.

Statistical analysis. Pearson's correlations between IQ and cortical thickness were estimated at each cortical point. Each subject contributed only one scan to maintain independence of data, and efforts were made to ensure a wide age range was covered. Developmental effects were explored by dividing the sample equally into four age groups (called early childhood (age range 3.8–8.4 yr), late childhood (range 8.6–11.7 yr), adolescence (11.8–16.9 yr) and early adulthood (17–29 yr)). Correlations for each of 56 brain subregions were Z-transformed, and the difference between the Z scores for each age group, and its significance, was calculated. To correct for the large number of comparisons, a false discovery rate of 0.05 was applied²⁸. Gender effects were examined for the entire sample in a similar manner.

To exploit the longitudinal nature of our data set, we used linear mixed-model regression, as this technique permits the inclusion of multiple measurements per person, missing data, and irregular intervals between measurements, thereby increasing statistical power while controlling for within-individual variation²⁹. Polynomial models for age effects were compared throughout the cerebral cortex and a cubic model found to provide the best fit, with the exception of anterior temporal cortices where a linear model was appropriate. A cubic model was therefore used to model age effects in the analyses presented. We first examined whether the relationship between IQ and cortical thickness differs with age by regressing cortical thickness at every vertex against IQ, age terms, and the interaction of IQ and age terms. For further exploration of the interaction, we divided the subjects into three IQ groups. This approach loses some power by categorizing a continuous variable, but has the advantage of rendering the results readily interpretable, allowing comparisons between highly intelligent and less intelligent groups. The resulting statistical maps were thresholded to control for multiple comparisons using the false discovery rate (FDR) procedure with $q = 0.05$ (refs 28, 30). An FDR threshold was determined for the statistical model using all *P* values pooled across all effects included in the model. At every cortical point, *t*-statistics were visualized through projection onto a standard brain template (the map shows the results of the interaction between the cubic age term and IQ groups). Such visualization showed clusters of cortical points that had a significant difference between the intelligence groups in the trajectory of cortical growth. The longitudinal analyses selected and averaged all cortical points within each of these clusters. Graphs illustrating the trajectories were generated using fixed-effects parameter estimates.

To illustrate differences in cortical thickness between the superior and average intelligence groups at different ages, linear mixed-models were run at different centred ages. For example, for age seven years, seven was subtracted from the age at scan acquisition, and this value entered as the age term. *t*-statistics representing the differences in cortical thickness between the two intelligence groups at each age were projected onto brain templates. This analysis represents group differences at each age based on values estimated from developmental curves modelled on all data.

Received 25 October; accepted 29 November 2005.

1. Spearman, C. 'General intelligence' objectively determined and measured. *Am. J. Psychol.* **15**, 201–293 (1904).
2. Gottfredson, L. S. Why *g* matters: The complexity of everyday life. *Intelligence* **24**, 79–132 (1997).
3. Deary, I. J., Whalley, L. J., Lemmon, H., Crawford, J. R. & Starr, J. M. The stability of individual differences in mental ability from childhood to old age: Follow-up of the 1932 Scottish Mental Survey. *Intelligence* **28**, 49–55 (2000).
4. Booth, J. R. et al. Neural development of selective attention and response inhibition. *Neuroimage* **20**, 737–751 (2003).
5. Sowell, E. R. et al. Longitudinal mapping of cortical thickness and brain growth in normal children. *J. Neurosci.* **24**, 8223–8231 (2004).
6. McDaniel, M. Big-brained people are smarter. *Intelligence* **33**, 337–346 (2005).
7. Wilke, M., Sohn, J. H., Byars, A. W. & Holland, S. K. Bright spots: correlations of gray matter volume with IQ in a normal pediatric population. *Neuroimage* **20**, 202–215 (2003).
8. Frangou, S., Chitins, X. & Williams, S. C. Mapping IQ and gray matter density in healthy young people. *Neuroimage* **23**, 800–805 (2004).
9. Haier, R. J., Jung, R. E., Yeo, R. A., Head, K. & Alkire, M. T. Structural brain variation and general intelligence. *Neuroimage* **23**, 425–433 (2004).
10. Kraemer, H. C., Yesavage, J. A., Taylor, J. L. & Kupfer, D. How can we learn about developmental processes from cross-sectional studies, or can we? *Am. J. Psychiatry* **157**, 163–171 (2000).
11. Giedd, J. N. et al. Brain development during childhood and adolescence: a longitudinal MRI study. *Nature Neurosci.* **2**, 861–863 (1999).
12. Wechsler, D. *Manual for the Wechsler Intelligence Scale for Children—Revised* (The Psychological Corporation, New York, 1974).
13. O'Donnell, S., Noseworthy, M. D., Levine, B. & Dennis, M. Cortical thickness of the frontopolar area in typically developing children and adolescents. *Neuroimage* **24**, 948–954 (2005).
14. Kabani, N., Le Goualher, G., MacDonald, D. & Evans, A. C. Measurement of cortical thickness using an automated 3-D algorithm: a validation study. *Neuroimage* **13**, 375–380 (2001).
15. Lerch, J. P. & Evans, A. C. Cortical thickness analysis examined through power analysis and a population simulation. *Neuroimage* **24**, 163–173 (2005).
16. Gray, J. R., Chabris, C. F. & Braver, T. S. Neural mechanisms of general fluid intelligence. *Nature Neurosci.* **6**, 316–322 (2003).
17. Duncan, J. et al. A neural basis for general intelligence. *Science* **289**, 457–460 (2000).
18. Kostovic, I., Judas, M., Rados, M. & Hrabac, P. Laminar organization of the human fetal cerebrum revealed by histochemical markers and magnetic resonance imaging. *Cereb. Cortex* **12**, 536–544 (2002).
19. Kostovic, I. & Rakic, P. Developmental history of the transient subplate zone in the visual and somatosensory cortex of the macaque monkey and human brain. *J. Comp. Neurol.* **297**, 441–470 (1990).
20. Yakovlev, P. I. & Lecours, A. R. in *Regional Development of the Brain in Early Life* (ed. Minokowski, A.) (Blackwell Scientific, Oxford, 1967).
21. Huttenlocher, P. R. & Dabholkar, A. S. Regional differences in synaptogenesis in human cerebral cortex. *J. Comp. Neurol.* **387**, 167–178 (1997).
22. Hensch, T. K. Critical period regulation. *Annu. Rev. Neurosci.* **27**, 549–579 (2004).
23. Chugani, H. T., Phelps, M. E. & Mazziotta, J. C. Positron emission tomography study of human brain functional development. *Ann. Neurol.* **22**, 487–497 (1987).
24. Collins, D. L., Neelin, P., Peters, T. M. & Evans, A. C. Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. *J. Comput. Assist. Tomogr.* **18**, 192–205 (1994).
25. Sled, J. G., Zijdenbos, A. P. & Evans, A. C. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans. Med. Imaging* **17**, 87–97 (1998).
26. Zijdenbos, A. P., Forghani, R. & Evans, A. C. Automatic "pipeline" analysis of 3-D MRI data for clinical trials: application to multiple sclerosis. *IEEE Trans. Med. Imaging* **21**, 1280–1291 (2002).
27. MacDonald, D., Kabani, N., Avis, D. & Evans, A. C. Automated 3-D extraction of inner and outer surfaces of cerebral cortex from MRI. *Neuroimage* **12**, 340–356 (2000).
28. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).
29. Pinheiro, J. C. & Bates, D. M. *Mixed-effects Models in S and S-PLUS* (Springer, New York, 2000).
30. Genovese, C. R., Lazar, N. A. & Nichols, T. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage* **15**, 870–878 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This research was supported by the Intramural Research Program of the National Institutes of Health. We acknowledge the statistical advice of G. Chen and technical assistance from T. Nugent III. The authors thank the children who participated in the study and their families.

Author Contributions P.S. designed and wrote the study with J.R. and J.G., and conducted neuroimaging analyses. J.G. and J.R. directed the project. D.G. conducted longitudinal analyses. L.C. was data manager, and R.L. and N.G. advised on interpretation and analysis. J.L. and A.E. developed cortical thickness analytic tools and J.L. developed software for longitudinal neuroimaging analyses.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.S. (shawp@mail.nih.gov).

LETTERS

Reverse replay of behavioural sequences in hippocampal place cells during the awake state

David J. Foster¹ & Matthew A. Wilson¹

The hippocampus has long been known to be involved in spatial navigational learning in rodents^{1,2}, and in memory for events in rodents^{3,4}, primates⁵ and humans⁶. A unifying property of both navigation and event memory is a requirement for dealing with temporally sequenced information. Reactivation of temporally sequenced memories for previous behavioural experiences has been reported in sleep in rats^{7,8}. Here we report that sequential replay occurs in the rat hippocampus during awake periods immediately after spatial experience. This replay has a unique form, in which recent episodes of spatial experience are replayed in a temporally reversed order. This replay is suggestive of a role in the evaluation of event sequences in the manner of reinforcement learning models. We propose that such replay might constitute a general mechanism of learning and memory.

We used multiple single-unit recording techniques⁹ to measure hippocampal neural activity during periods of running and stopping in four rats. Two sessions were recorded per animal, one on a familiar track and one on a new track. During each session, the animal ran several laps, with each lap consisting of running from one end of the track to the other and back again. Within a given lap, the animal stopped at each end to consume food from a food well. After consuming the food, the animal would wait of its own accord in the same position for a short period of time that varied from lap to lap (Fig. 1a). The behaviour of the animal during this time varied between grooming, whisking or being still. The animal would then turn around and immediately begin running again.

For each recording session, we first characterized the activity of neurons in terms of their place fields¹⁰ during locomotion, as measured using the spikes from all laps (with each running direction considered separately; Fig. 1a). Neurons satisfying minimum firing rate and waveform criteria were selected (see Methods), and their place fields were ordered according to the position of the field peaks (Fig. 1c) in order to generate a probe sequence. This probe sequence was then used to examine patterns of activity in cells during individual laps (Fig. 2a, b). While an animal was running, cells fired in order with respect to position, as expected from their place fields. However, during the stopping periods immediately after running, regularly occurring instances of coincident spiking were evident, involving many of the cells in the probe sequence. Notably, within each coincident event, the sequence of cell activation was in reverse order with respect to the probe sequence, and spanned the equivalent of the entire track, on a timescale of hundreds of milliseconds (Fig. 2c).

To quantify the effect for each recording session, we first identified coincident spiking events during stopping periods that involved a large proportion of the cells in the probe sequence for that recording session (see Methods). For each event, the rank-order correlation between cell number and time was calculated, together with a probability¹¹. Examples of significant ($P < 0.05$) events are shown

in Fig. 3 (Supplementary Figs S2–S5 show example events for all four animals). For each of the eight sessions, over both directions, the distribution of correlation values of all events (regardless of P value) was found to be significantly different from (that is, significantly negative with respect to) the distribution of correlation values of all events with the cell-order parameter shuffled randomly (Fig. 4; P values in figure legend). Hence, the occurrence of reverse replay events was significantly greater than would be expected by chance. The correlation distribution of all events across all four new sessions was significantly different from (that is, more negative than) the distribution of all events across all four familiar sessions (two-tailed Kolmogorov–Smirnov test, $P = 1.13 \times 10^{-10}$), indicating that the phenomenon is more readily observable in a new environment. A number of cells were bidirectional, in that they did not have a peak firing rate in a preferred direction that was at least double that in the opposite direction (52% bidirectional neurons in the new sessions; 35% in the familiar sessions), raising the possibility that apparently reverse replay events merely reflected forward replay of neurons in

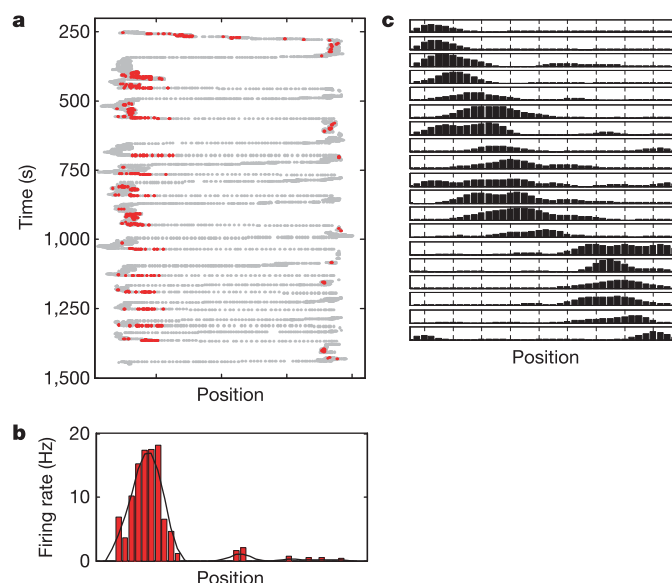
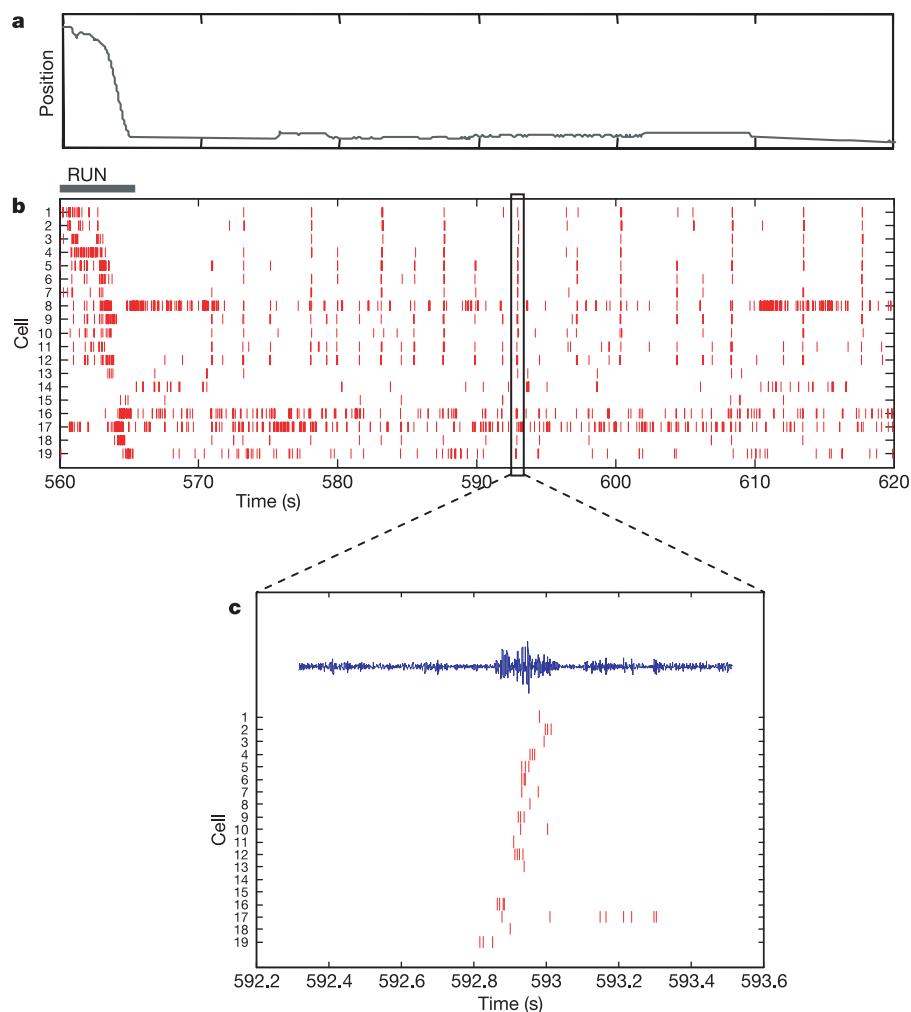


Figure 1 | A sequence of place fields. **a**, The position of a rat during one recording session is shown in grey as a function of time. Stopping periods at each end could exceed one minute. Spikes emitted by a single hippocampal place cell while the animal faced rightwards are shown in red. **b**, The place field of the single cell shown in Fig. 2a. **c**, Simultaneous recording of 128 cells, of which 26 cells had place fields on the track. Nineteen cells with fields in the rightward direction were ordered by peak to generate a probe sequence.

¹Picower Institute for Learning and Memory, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 77 Massachusetts Avenue, 46-5223, Cambridge, Massachusetts 02139, USA.

Figure 2 | Reverse replay events during a single lap. **a**, The position of the rat as function of time. **b**, Spikes are shown for each of the cells from Fig. 1c, in the same sequential order. The x axis (time) is the same as, and aligned with, panel **a**. During the stopping period, coincident spiking events are visible as narrow, vertical lines. **c**, A section from **b** with the x axis expanded to reveal reverse replay. In blue, the simultaneously recorded hippocampal EEG shows a co-occurring ripple event.



the opposite direction. Each of 486 reverse replay events was assessed with the bidirectional cells removed, of which 117 remained significant ($P < 10^{-13}$ under a binomial distribution). Only 8 of the 486 corresponding simultaneous events in the other direction remained significant with the bidirectional cells removed ($P = 1$). Hence, unidirectional cells from the probe sequences for the preferred direction showed significant reverse replay, whereas unidirectional cells for the opposite direction did not show significant forward replay.

Most stopping periods with reverse replay showed multiple reverse replay events (Fig. 4b). Reverse replay occurred even after the first lap on a new track (Fig. 3). Reverse replay events were coincident with ripples in the hippocampal electroencephalogram (EEG; Fig. 4c), which are characteristic of hippocampal activity during both awake, non-running periods and sleep^{12–14}. The question remained as to whether reverse replay reflected immediate experience, and so memory for the experiential sequence, or whether the replay could occur in the absence of immediate experience, reflecting some pre-existing expectation of sequential order. In six sessions, we recorded cell activity after the animal had been placed on the track but before running, during which time the animal was still and facing away from the track, hence in a similar physical state to that occupied during subsequent stopping periods in that location. None of these periods showed reverse replay, although the periods ranged between 42.3 s and 424.3 s in duration. A possible model for the generation of reverse replay sequences that encompasses these data is presented in Supplementary Fig. S6.

The hippocampus has long been known to be necessary for learning in sequential decision problems such as navigation^{1–3}.

Sequential decision problems suffer from the well-known temporal credit assignment problem—that of relating reward information that might occur only at the end of a sequence of events to the individual events within that sequence. A classic solution to this problem is to propagate value information from the rewarded location backwards along incoming trajectories^{15–22}. In the brain, reverse replay could be paired with a fast-onset, slowly decaying dopamine signal to learn a representation of value, thus providing a value gradient that the animal could follow during subsequent goal-finding behaviour (Supplementary Fig. S7). Hence, reverse replay in the hippocampus might have a critical role in support of learning in hippocampus-dependent tasks. The finding that reverse replay is more readily observable in a new environment than a familiar one is consistent with such a role.

Reverse replay during the awake state can be contrasted with replay in sharp waves during slow-wave sleep, in which episodes of spatial experience are replayed in the same temporal order as that in which they were experienced⁸. This re-expression of events while the animal occupies an entirely different physical and temporal context, as well as a different behavioural state, may have a role in memory consolidation during sleep^{23,24}. When awake, reverse replay occurs *in situ*, allowing immediately preceding events to be evaluated in precise temporal relation to a current, anchoring event, and so may be an integral mechanism for learning about recent events. Moreover, by converting single experiences into multiple reverse events, even after the first encounter in a new environment, awake replay represents efficient use of hard-won experience. Understanding this replay is likely to be critical to understanding how animals learn from experience.

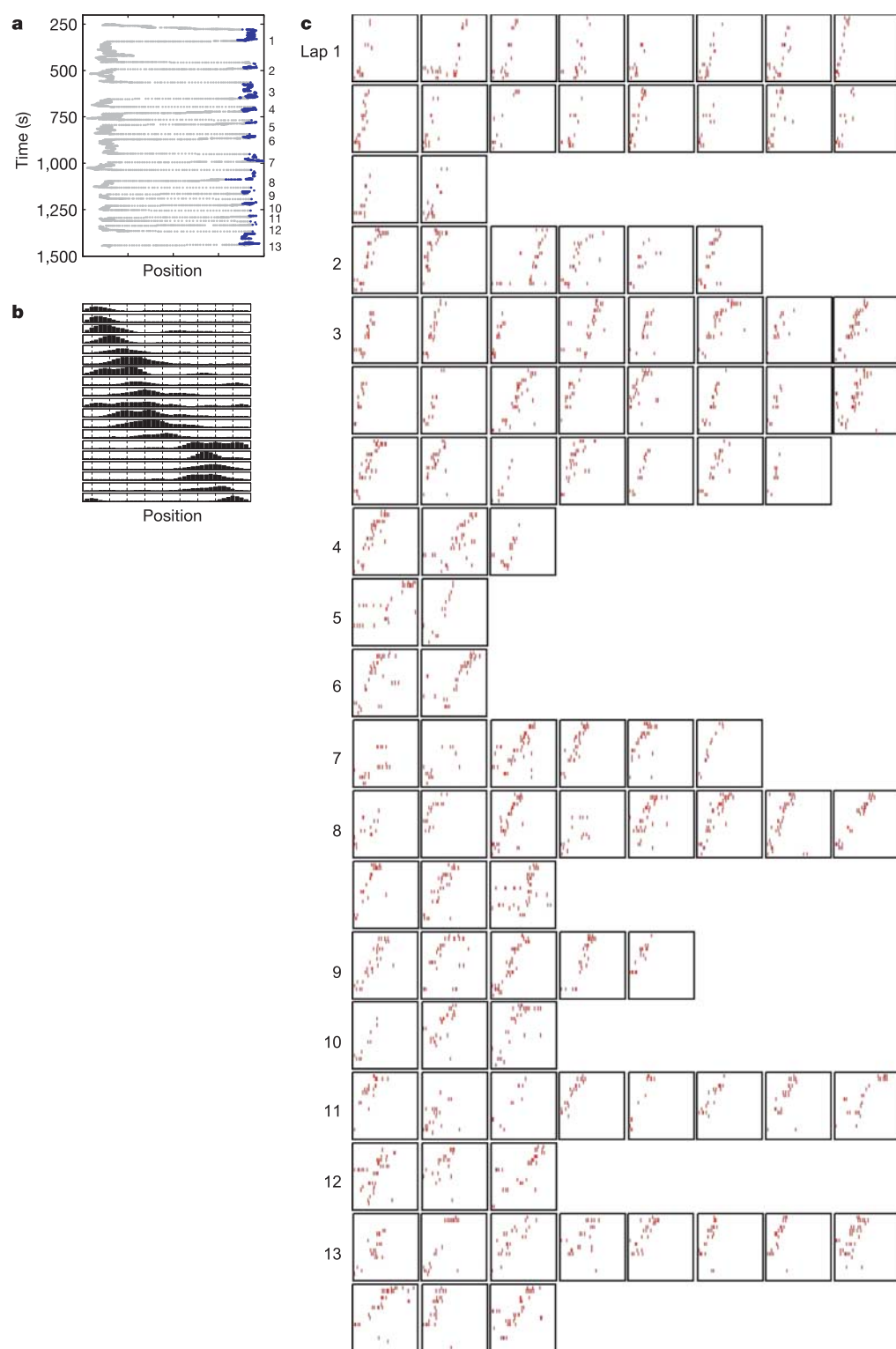


Figure 3 | Reverse replay events, by lap, for an entire recording session. **a**, The session follows rat 2 on the new track. Position during running periods is shown in grey; stopping periods (for the rightward direction only) are shown in blue. Laps are numbered on the right. **b**, The probe sequence for the rightward direction, showing the position of the 19 place fields in the sequence. **c**, Reverse replay events by lap. The y axis in each case is cells 1–19. The x axis in each case is a fixed time-window of 288 ms.

METHODS

Electrophysiology and behavioural apparatus. In each of four rats, a multiple electrode microdrive array⁹ consisting of either 18 (rats 1 and 4) or 17 (rats 2 and 3) independently adjustable tetrodes was implanted above the right dorsal hippocampus (4 mm posterior, 2.2 mm lateral with respect to bregma), and the tetrodes were lowered over the course of several days until they rested in the CA1 pyramidal cell layer. The remaining details of the procedure were as previously described⁸. Direction was measured by the relative position of two tracker diodes mounted to the front and rear of the tetrode drive. Linear tracks (162 cm long) were used for both new and familiar sessions for rats 1–3, and a U-shaped track (205 cm long, 45 cm wide) was used for both new and familiar sessions for rat 4.

Place-field analysis. Position was linearized for each session to yield a scalar

value of distance along the track. A histogram of spikes from each cell was calculated over position bins and was normalized by the time spent by the animal in each bin, to yield a place field. Fields were velocity-filtered to exclude times when the speed of the animal was below 5.4 cm s^{-1} . In order to assign a peak value, the histogram was smoothed (as shown by the black line in Fig. 2b). Cells with a peak firing rate of at least 5 Hz were included in the probe sequence, with the exception of putative inhibitory interneurons, which were identified as cells with a mean peak-to-trough spike width of less than 0.35 ms.

Spike-train analysis. A spike train was constituted from all spikes (from all cells in the probe sequence) that occurred during stopping periods while the animal faced in the direction in which it had just run. This spike train was then broken between every pair of successive spikes separated by more than 50 ms, to form a

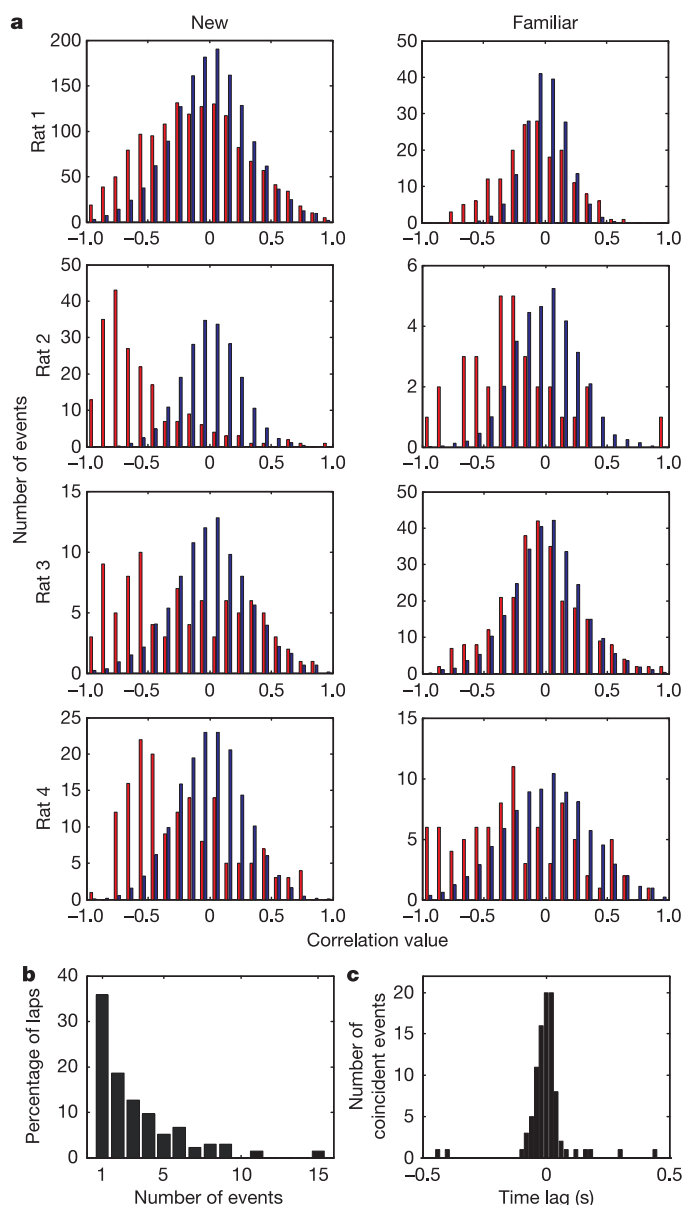


Figure 4 | Analysis of reverse replay across all recording sessions. **a**, For each session, a histogram of the rank-order correlation values of all events is shown in red, and a histogram of shuffled correlation values is shown in blue (see Methods). The two distributions were statistically different. *P* values for each session were as follows, where *n* is the total number of events. New: Rat 1, *n* = 1,425, *P* = 3.04×10^{-41} ; Rat 2, *n* = 202, *P* = 4.00×10^{-98} ; Rat 3, *n* = 91, *P* = 6.52×10^{-10} ; Rat 4, *n* = 160, *P* = 4.88×10^{-21} . Familiar: Rat 1, *n* = 178, *P* = 6.69×10^{-8} ; Rat 2, *n* = 33, *P* = 4.05×10^{-6} ; Rat 3, *n* = 275, *P* = 0.0067; Rat 4, *n* = 88, *P* = 1.32×10^{-8} . The percentage of events with significant reverse correlations was as follows, by session: New, Rat 1, 13%; Rat 2, 72%; Rat 3, 31%; Rat 4, 29%; Familiar, Rat 1, 19%; Rat 2, 30%; Rat 3, 6%; Rat 4, 16%. **b**, Histogram of the number of significant reverse events per stopping period, for those stopping periods with at least one significant event. **c**, Cross-correlogram of significant reverse replay events with hippocampal sharp waves, for an example session in which there were 94 coincident events out of a total of 146 replay events.

large set of proto-events. Those proto-events in which at least one-third of the cells in the probe sequence fired at least one spike were then selected as events. The few events longer than 500 ms in duration were rejected as a potential source of spurious correlations. For each event, 100 shuffled events were created by randomly permuting the cell-order parameter. The histograms in Fig. 4a were normalized (by dividing by 100) to allow visual comparison with the original distributions. A non-parametric, two-sample Kolmogorov–Smirnov test was used to determine whether the distributions were significantly different.

Ripple identification. Sharp waves reverse at about the electrode depth corresponding to maximum cell yield in the hippocampus, making it difficult to measure sharp waves directly. However, they co-occur with transient, high-frequency events called ripples (100–400 Hz). Ripples were identified as reported previously⁸. A single time of occurrence for each ripple was calculated as the mean of the start and end times of the ripple. A single time was similarly calculated for each replay event. These times were used to generate a cross-correlogram, which was not normalized, so that the *y* axis of Fig. 4c is in numbers of coincident events. Values for total numbers of coincident events cited in the text were found by summing the values of bins between –50 ms and 50 ms.

Received 13 December 2005; accepted 20 January 2006.

Published online 12 February 2006.

- O'Keefe, J. & Nadel, L. *The Hippocampus as a Cognitive Map* (Clarendon, London, 1978).
- Morris, R. G. M., Garrud, P., Rawlins, J. N. P. & O'Keefe, J. Place navigation impaired in rats with hippocampal lesions. *Nature* **297**, 681–683 (1982).
- Olton, D. S. & Samuelson, R. J. Remembrance of places past: spatial memory in rats. *J. Exp. Psychol. Anim. Behav. Process.* **2**, 97–116 (1976).
- Steele, R. J. & Morris, R. G. M. Delay-dependent impairment of a matching-to-place task with chronic and intrahippocampal infusion of the NMDA-antagonist D-AP5. *Hippocampus* **9**, 118–136 (1999).
- Gaffan, D. Scene-specific memory for objects: a model of episodic memory impairment in monkeys with fornix transection. *J. Cogn. Neurosci.* **6**, 305–320 (1994).
- Scoville, W. B. & Milner, B. Loss of recent memory after bilateral hippocampal lesions. *J. Neurol. Neurosurg. Psychiatry* **20**, 11–21 (1957).
- Louie, K. & Wilson, M. A. Temporally structured replay of awake hippocampal ensemble activity during rapid eye movement sleep. *Neuron* **29**, 145–156 (2001).
- Lee, A. K. & Wilson, M. A. Memory of sequential experience in the hippocampus during slow wave sleep. *Neuron* **36**, 1183–1194 (2002).
- Wilson, M. A. & McNaughton, B. L. Dynamics of the hippocampal ensemble code for space. *Science* **261**, 1055–1058 (1993).
- O'Keefe, J. & Dostrovsky, J. The hippocampus as a spatial map: preliminary evidence from unit activity in the freely moving rat. *Brain Res.* **34**, 171–175 (1971).
- Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. *Numerical Recipes in C: The Art of Scientific Computing* 2nd edn (Cambridge Univ. Press, Cambridge, 1992).
- Ranck, J. B. Jr. Studies on single neurons in dorsal hippocampal formation and in septum in unrestrained rats. I. Behavioral correlates and firing repertoires. *Exp. Neurol.* **42**, 461–531 (1973).
- Buzsáki, G. Hippocampal sharp waves: their origin and significance. *Brain Res.* **398**, 242–252 (1986).
- Buzsáki, G. A two-stage model of memory trace formation: a role for 'noisy' brain states. *Neuroscience* **31**, 551–570 (1989).
- Kaelbling, L. P., Littman, M. L. & Moore, A. W. Reinforcement learning: a survey. *J. Artif. Intell. Res.* **4**, 237–285 (1996).
- Sutton, R. S. Learning to predict by the methods of temporal differences. *Mach. Learn.* **3**, 9–44 (1988).
- Schultz, W., Dayan, P. & Montague, P. R. A neural substrate of prediction and reward. *Science* **275**, 1593–1599 (1997).
- Foster, D. J., Morris, R. G. M. & Dayan, P. A model of hippocampally dependent navigation, using the temporal difference learning rule. *Hippocampus* **10**, 1–16 (2000).
- Lin, L. J. Self-improving reactive agents based on reinforcement learning, planning and teaching. *Mach. Learn.* **8**, 293–321 (1992).
- Sutton, R. S. & Singh, S. P. in *Proceedings of the Eighth Yale Workshop on Adaptive and Learning Systems* 31–36 (Yale Univ., New Haven, 1994).
- Cichosz, P. An analysis of experience replay in temporal difference learning. *Cybernet. Syst.* **30**, 341–363 (1999).
- Reynolds, S. *Experience Stack Reinforcement Learning for Off-policy Control*. Cognitive Science Technical Report CSR02-1 (Birmingham Univ., Birmingham, 2002).
- Stickgold, R., Whidbee, D., Schirmer, B., Patel, V. & Hobson, J. A. Visual discrimination task improvement: a multi-step process occurring during sleep. *J. Cogn. Neurosci.* **12**, 246–254 (2000).
- Siapas, A. G. & Wilson, M. A. Coordinated interactions between hippocampal ripples and cortical spindles during slow-wave sleep. *Neuron* **21**, 1123–1128 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank V. Ego-Stengel, P. Dayan and M. Fee for comments on the manuscript. This work was supported by an NIH grant to M.A.W.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.J.F. (djfofster@mit.edu).

LETTERS

A *C. elegans* stretch receptor neuron revealed by a mechanosensitive TRP channel homologue

Wei Li¹, Zhaoyang Feng¹, Paul W. Sternberg³ & X. Z. Shawn Xu^{1,2}

The nematode *Caenorhabditis elegans* is commonly used as a genetic model organism for dissecting integration of the sensory and motor systems¹. Despite extensive genetic and behavioural analyses that have led to the identification of many genes and neural circuits involved in regulating *C. elegans* locomotion behaviour¹, it remains unclear whether and how somatosensory feedback modulates motor output during locomotion. In particular, no stretch receptors have been identified in *C. elegans*, raising the issue of whether stretch-receptor-mediated proprioception is used by *C. elegans* to regulate its locomotion behaviour. Here we have characterized TRP-4, the *C. elegans* homologue of the mechanosensitive TRPN channel. We show that *trp-4* mutant worms bend their body abnormally, exhibiting a body posture distinct from that of wild-type worms during locomotion, suggesting that TRP-4 is involved in stretch-receptor-mediated proprioception. We show that TRP-4 acts in a single neuron, DVA, to mediate its function in proprioception, and that the activity of DVA can be stimulated by body stretch. DVA both positively and negatively modulates locomotion, providing a unique mechanism whereby a single neuron can fine-tune motor activity. Thus, DVA represents a stretch receptor neuron that regulates sensory-motor integration during *C. elegans* locomotion.

Transient receptor potential (TRP) proteins represent a superfamily of cation channels that are conserved from worms to humans and comprise seven subfamilies (TRPC, TRPV, TRPM, TRPN, TRPA, TRPP and TRPML)². TRP channels have been implicated in various physiological processes ranging from fertilization to mechanosensation^{2,3}. We are particularly interested in TRP-4, a TRPN channel, because of its potential role in regulating mechanosensation (see below). We isolated two deletion mutants of TRP-4. Both *trp-4* alleles lack the regions encoding transmembrane domains and are likely to be null (Fig. 1a). TRP-4 has ~40% sequence identity to and shares similar domain structures with zebrafish TRPN1 and *Drosophila* NOMPC, which encode putative mechanosensitive channels required for detecting sound vibration by hair cells in zebrafish and for sensing bristle displacement in flies, respectively^{4,5}.

During locomotion, worms bend their body periodically, propagating a sinusoidal wave along their body axis^{1,6}. We used an automated worm tracking system⁷ to record worm locomotion (see Methods), as conventional methods (human description) cannot provide quantitative measurement. Digitized images were then processed and subjected to data analysis. To facilitate data processing, we divided the worm body into 12 segments such that various locomotion parameters could be readily calculated, including frequency of body bending, extent of body bending (bending angles), track amplitude and track wavelength⁷ (Fig. 1b).

Both *trp-4* alleles showed two distinct locomotion defects. *trp-4* worms bent their body more frequently (Fig. 1e), a phenotype that we named 'fast bending'. As a result, the centroid speed in the mutant

worms was increased (Fig. 1g). In addition, these mutant worms bent their body more deeply and showed a body posture distinct from that of wild-type worms during locomotion (Fig. 1c, d, f), suggesting a defect in stretch-receptor-mediated proprioception. This observation is consistent with the role of TRPN channels in mechanosensation, because proprioception is mediated by mechanosensitive channels^{4,5,8}. As a result of this locomotion defect, mutant worms left behind deeper sinusoidal tracks than did wild-type worms (Fig. 1c, d). We named this second *trp-4* phenotype the 'exaggerated bending' phenotype. We used the extent of body bending (bending angles) to describe the exaggerated bending phenotype, because this parameter alone readily quantifies the curvature of the worm body independently of the body length. Neither the track amplitude nor the track wavelength alone is sufficient to do so. Similar methods are used in the clinic to diagnose scoliosis⁹. Both *trp-4* defects were

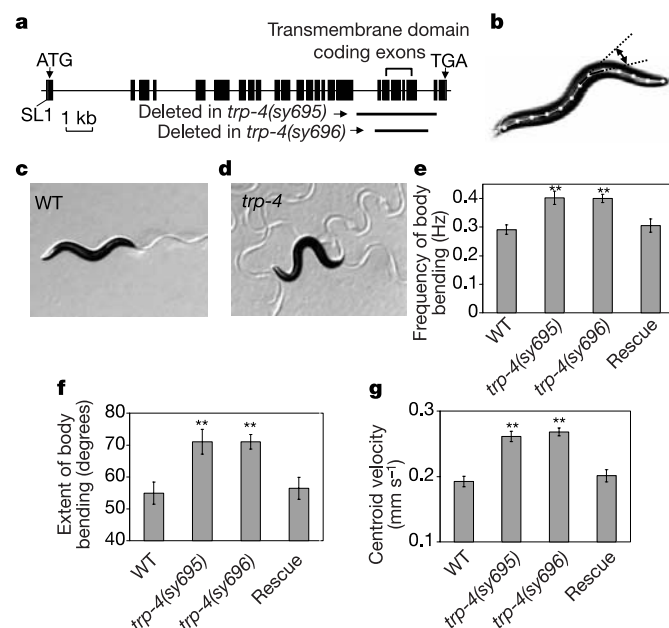


Figure 1 | Defective locomotion in *trp-4* mutants. **a**, *trp-4* gene structure and *trp-4* mutations. **b**, Representative image showing that the worm body is divided into 12 segments (adapted from ref. 7). The arrow between the two dotted lines denotes the extent of body bending (bending angle) between the two segments. **c**, **d**, Snapshot images of a moving wild-type worm (**c**) and a *trp-4(sy695)* worm with abnormal body posture (**d**). **e**, Increased frequency of body bending in *trp-4* mutants. **f**, *trp-4* mutant worms bend their body more deeply. **g**, Centroid velocity is increased in *trp-4* mutants. Rescue indicates *trp-4(sy695)* mutants expressing *Ex[trp-4::yfp]*. Error bars indicate s.e.m. ($n \geq 12$). ** $P < 0.005$.

¹Life Sciences Institute, University of Michigan, and ²Department of Molecular and Integrative Physiology, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA. ³Howard Hughes Medical Institute and Division of Biology, California Institute of Technology, Pasadena, California 91125, USA.

rescued by a transgene encoding full-length TRP-4 fused to yellow fluorescent protein (Fig. 1e–g).

trp-4 has been reported to be expressed in the CEP and ADE dopamine neurons and in two interneurons, DVA and DVC⁵. We also observed TRP-4 expression in the PDE dopamine neurons, probably owing to our use of a longer promoter region (~7.5 kb) of the *trp-4* gene. TRP-4 was highly enriched in the cilia of the dopamine neurons (Supplementary Fig. S1a), and was localized throughout the whole axon in DVA and DVC⁵ (Supplementary Fig. S1b).

C. elegans dopamine neurons are sensory neurons with a morphology analogous to that of vertebrate hair cells in the inner ear⁶ (Supplementary Fig. S1c). Unlike the typical touch receptor neurons that detect gentle touch^{10,11}, these dopamine neurons are mechanosensory neurons that sense mechanical attributes imposed by the surface material on which worms navigate¹². As a result, worms slow down their frequency of body bending after encountering bacteria, a phenomenon called the ‘basal slowing response’¹². In the absence of bacteria, the frequency of body bending in *trp-4* mutant worms was no longer faster than that in wild-type worms, suggesting that the fast bending phenotype of *trp-4* mutants might be due to a defect in the basal slowing response (Fig. 2a). Consequently, *trp-4* mutant worms might always be in the highest state of locomotion, which in wild-type worms occurs only in the absence of bacteria¹². In support of this idea, the dopamine-deficient mutant *cat-2(e1112)* (ref. 13) showed the same fast bending phenotype as the *trp-4* mutant worms (Fig. 2a). In addition, expression of wild-type copies of *trp-4* specifically in dopamine neurons rescued the fast bending phenotype in *trp-4* mutants (Fig. 2a). Thus, dopamine neurons seem to mediate the fast bending phenotype in *trp-4* mutants.

The extent of body bending in wild-type and *trp-4* mutant worms was not affected by the presence of bacteria (Fig. 2b), however, suggesting that dopamine neurons are unlikely to mediate the exaggerated bending phenotype. Consistent with this hypothesis, the extent of body bending in the dopamine-deficient mutant *cat-2* was similar to that of wild type (data not shown). We therefore examined the involvement of DVA and DVC, which also express TRP-4. These two neurons have their somata situated in the worm tail, and their axons span nearly the whole length of the worm body (Fig. 3a and Supplementary Fig. S1b). As TRP-4 is localized throughout the axons of DVA and DVC (Supplementary Fig. S1b), we speculated that when a worm bends its body, it stretches the plasma membrane of the axons of DVA and DVC. This stretch may then lead to activation of TRP-4, a mechanosensitive channel homologue, and consequently stimulate DVA and DVC (Fig. 3a). These two neurons would then signal negatively to the downstream command interneurons and ventral cord motor neurons onto which they primarily

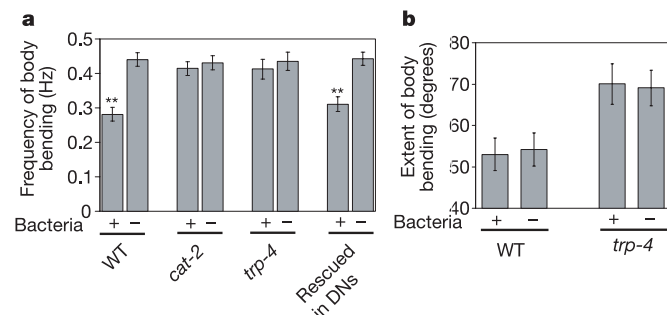


Figure 2 | Dopamine neurons mediate the fast bending phenotype of *trp-4* mutants. **a**, The fast bending phenotype of *trp-4* mutants is mediated by dopamine neurons. Worms were tracked on plates with or without bacteria. DNs, dopamine neurons. Rescue indicates *trp-4(sy695)* mutants expressing *Ex[Pdat-1::trp-4]*. **b**, Extent of body bending is not affected by the presence of bacteria in wild-type or *trp-4(sy695)* worms. Data were processed from the same samples as in **a**. Error bars indicate s.e.m. ($n \geq 10$). $^{**}P < 0.005$.

synapse¹⁴, inhibiting body-wall muscle contraction and thereby preventing exaggerated body bending. The command interneurons and ventral cord motor neurons are the two key components in the locomotion circuitry⁶. This model predicts that laser ablation of DVA and DVC in wild-type worms should mimic the exaggerated body bending seen in *trp-4* mutant worms. Instead of augmenting the extent of body bending, however, laser ablation of DVA and DVC in wild-type worms slightly reduced the extent of body bending (Fig. 3e). Although the above proposed model seemed to be incorrect, this observation showed that DVA and DVC are important in regulating the extent of body bending.

Because killing of DVA and DVC led to a reduction in the extent of body bending, we considered that there might be a positive regulator of the extent of body bending in these two neurons in addition to the negative regulator TRP-4 (Fig. 3b). If so, laser ablation of DVA and DVC in wild-type worms would not be expected to recapitulate the *trp-4* phenotype, because such ablation would eliminate both negative and positive regulators. This second model would also explain the exaggerated bending phenotype of *trp-4* mutants, because loss of the negative regulator TRP-4 would unmask the effect of the putative positive regulator. If this second model is correct, laser ablation of DVA and DVC in the *trp-4* mutant background should abrogate the activity of the remaining positive regulator, and should hence suppress the *trp-4* phenotype. Indeed, laser ablation of the DVA neuron alone was sufficient to suppress the exaggerated body bending phenotype in *trp-4* mutant worms (Fig. 3c–e), whereas killing of DVC did not result in a significant effect (Fig. 3e). Expression of wild-type copies of TRP-4 specifically in DVA was also sufficient to rescue the exaggerated bending phenotype (Fig. 3e). These results suggest that the exaggerated bending phenotype of *trp-4* mutants is mediated by the DVA neuron. Given that body bending periodically exerts local stretch on the plasma membrane of the DVA axon (Fig. 3a), our data suggest that DVA might function as a stretch-sensitive neuron.

To provide physiological evidence that the DVA neuron is stretch-sensitive, we engineered a transgenic line expressing the genetically encoded Ca^{2+} sensor G-CaMP in the DVA neuron. DsRed2 was coexpressed with G-CaMP in DVA as an internal reference marker. G-CaMP has been successfully used as a non-invasive Ca^{2+} sensor in *C. elegans* and *Drosophila* neurons^{15,16}. We first immobilized the DVA

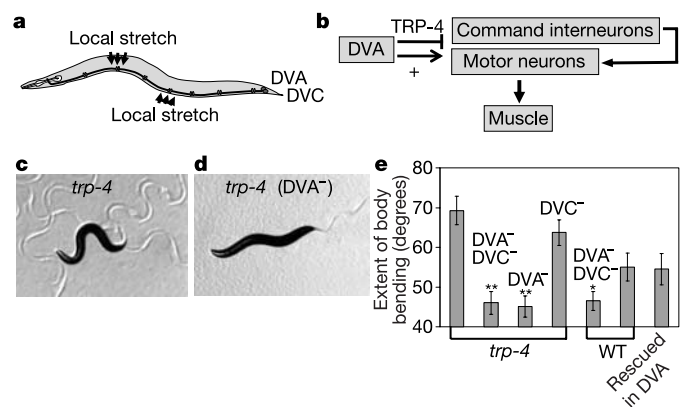


Figure 3 | TRP-4 functions in DVA to regulate the extent of body bending. **a**, Model showing that body bending locally stretches the plasma membrane of the DVA axon, potentially activating TRP-4 channels. Ovals depict TRP-4; arrows indicate the local stretch resulting from body bending. **b**, Second model proposing that the negative regulator TRP-4 acts together with an unknown positive factor in DVA to modulate the extent of muscle contraction. **c**, **d**, Snapshot images of a moving *trp-4(sy695)* worm (**c**; duplicate of Fig. 1d) and a moving DVA-ablated *trp-4(sy695)* worm (**d**), showing that the exaggerated bending phenotype has been suppressed. **e**, DVA mediates the exaggerated bending phenotype in *trp-4* mutants. Rescue indicates *trp-4(sy695)* mutants expressing *Ex[Ptwk-16(DVA)::trp-4]*. Error bars indicate s.e.m. ($n \geq 8$). $^{**}P < 0.005$; $^{*}P < 0.05$.

soma by gluing the tail of the worm on an agarose pad, while leaving the rest of the body free to move. Under these conditions, worms usually showed little movement, and no or little change in Ca^{2+} concentration was observed in DVA (Fig. 4b). On application of solution to the pad, worms began to bend their body in the liquid, triggering a robust increase in Ca^{2+} level in DVA ($n = 29/29$; Fig. 4a–c, e). We often observed repetitive Ca^{2+} spikes, with each spike presumably representing a body bending event (Fig. 4b). For those worms that vigorously bent their body at high frequency (>3 Hz), no sharp Ca^{2+} spikes were observed (Fig. 4c), probably owing to the relatively slow dissociation kinetics of G-CaMP (half time = 200 ms)¹⁷. We did not detect a significant Ca^{2+} response in *trp-4* mutant worms under the same conditions ($n = 0/16$; Fig. 4d, e), apart from a very brief Ca^{2+} transient that was often observed at the onset of liquid application ($n = 8/16$; Fig. 4d). Such a deficit in Ca^{2+} response was unlikely to be due to a defect in DVA excitability, because DVA isolated from *trp-4* mutant embryos retained the ability to respond to membrane depolarization induced by potassium chloride (Supplementary Fig. S3). These data suggest that body stretch is sufficient to stimulate TRP-4-dependent activity in DVA.

As DVA receives synaptic input from other mechanosensory neurons including PDE, PLM and PVD, we carried out similar imaging experiments on PDE-ablated wild-type worms and *mec-3* mutant (lacking PLM and PVD) worms^{18,19}, and observed similar Ca^{2+} responses (Fig. 4e and Supplementary Fig. S2c–f). No significant difference was detected in the extent of body bending between these worms and wild-type worms (Fig. 4f). We also

observed body-bending-evoked Ca^{2+} transients in *unc-13* worms, in which synaptic transmission is essentially eliminated²⁰ (Fig. 4e and Supplementary Fig. S2g, h). Thus, DVA seems to be the primary neuron mediating the body-bending-evoked Ca^{2+} transients.

To obtain further evidence that DVA is stretch-sensitive, we immobilized the worm's tail with glue, held its nose tip with a glass pipette, and then manually bent its body (Fig. 4g). Bending the worm's body evoked a sustained increase in Ca^{2+} in the DVA neuron, which decayed to basal levels after cessation of the stimulus ($n = 11/12$; Fig. 4h). The amplitude of the response seemed to be graded (Supplementary Fig. S4), and no sustained response was observed until the bending angle reached a specific threshold ($\sim 50^\circ$), consistent with a role for TRP-4 in antagonizing overcontraction of body-wall muscles (Fig. 4h and Supplementary Fig. S4). No such response was detected in *trp-4* mutant worms ($n = 0/11$), although a very brief Ca^{2+} transient was observed ($n = 11$; Fig. 4i, j). These observations indicate that DVA may be stretch-sensitive.

In summary, we have presented evidence supporting the notion that DVA is a stretch receptor neuron. Our results indicate that stretch-receptor-mediated proprioception is important for proper motor function in *C. elegans*. Nevertheless, our study does not exclude the presence of additional stretch receptors in *C. elegans*, for example, the undifferentiated processes of ventral cord motor neurons⁶. We have also shown the presence of a putative positive regulator in DVA; however, its activity might not be primarily mediated by Ca^{2+} . We propose that such a positive regulator may function to promote the extent of muscle contraction initially, and

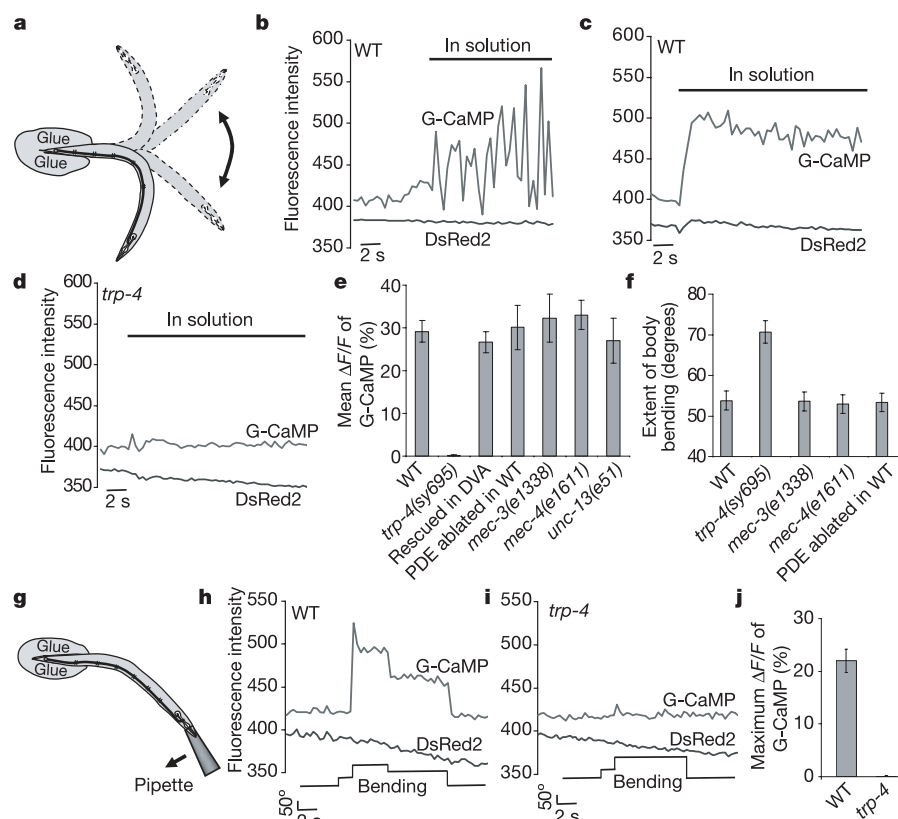


Figure 4 | The DVA neuron is stretch-sensitive. **a**, A worm, with its tail glued, freely bends its body. **b**, **c**, Body-bending-evoked Ca^{2+} signals in DVA. **d**, Body bending does not induce a significant Ca^{2+} response in *trp-4*(*sy695*). **e**, Mean G-CaMP fluorescence change in DVA. The mean peak fluorescence of G-CaMP in the first 20 s after liquid application was measured. Rescued in DVA indicates *trp-4*(*sy695*) mutants expressing *Ex[Ptrw-16(DVA)::trp-4]*. **f**, Lack of PDE or gentle- and harsh-touch receptor neurons does not have a significant effect on the extent of body bending. **g**, Manual bending of the

body of a glued worm by a glass pipette. **h**, **i**, Manual bending of a worm's body induces an increase in Ca^{2+} in DVA of wild-type (**h**), but not *trp-4*(*sy695*) mutant (**i**) worms. Representative traces are shown. **j**, Maximal increase in G-CaMP fluorescence in DVA. Wild-type and mutant worms were bent to $\sim 120^\circ$, at which point the Ca^{2+} concentration in DVA reached its maximum (Supplementary Fig. S5). The fluorescence intensity of DsRed2 slowly decreased because of its relatively fast bleach as compared with G-CaMP. Error bars indicate s.e.m. ($n \geq 5$).

that when muscle contraction reaches a specific extent, the negative regulator TRP-4 then signals to prevent further contraction of muscle cells. This dual control of the extent of body bending by DVA would confer on worms the capacity to tune the extent of body-wall muscle contraction, providing a unique mechanism for the general fine control of motor activity by proprioceptive stretch receptors.

Among the best-characterized stretch receptors in vertebrates are muscle spindles and Golgi tendon organs, whereas chordotonal organs represent the best-analysed proprioceptors in arthropods^{8,21}. In humans, muscle spindles signal to promote muscle contraction, and Golgi tendon organs function to repress muscle contraction to avoid muscle damage⁸. Such antagonistic roles of vertebrate muscle spindles and Golgi tendon organs seem analogous to those of the DVA neuron in *C. elegans*, except that DVA, as a single cell, seems to carry out both functions. Thus, the phenomenon that proprioceptor-mediated somatosensory feedback can both positively and negatively modulate muscle activity seems to be present in both organisms. We propose that some of the basic principles underlying somatosensory feedback regulation of motor output are evolutionarily conserved.

METHODS

Behavioural analysis and the worm tracker. L4 hermaphrodites were picked 16 h before behavioural analysis. Worms were tracked for 5 min at 20° on NGM plates spread with a thin layer of freshly grown OP50 bacteria as described⁷. For tracking in the absence of bacteria, the supernatant of OP50 culture was spread on tracking plates. The tracking system consists of a stereomicroscope mounted with a Cohu 7800 digital camera, a digital motion system (Parker Automation) that follows worm movement, and laboratory-developed software. The vision/motion data were compressed and integrated into AVI format for feature extraction. To quantify the extent of body bending (bending angles), binarized worm images were thinned to obtain the 'skeleton image' of the worm and broken into 12 equal-length segments with real world coordinates. For simplicity and consistency, we selected the middle segments (segments 6 and 7) of the worm body for quantification unless otherwise specified. Our wild-type and *cat-2* data are quantitatively similar to reported data¹², although our frequency of body bending is a little slower (~20%), probably because we averaged data from the whole tracking period (5 min). The centroid velocity was calculated by a described method⁷, which measures the speed of wave propagation but not the vector speed during locomotion.

Molecular biology. The 5' and 3' ends of the *trp-4* coding regions (Fig. 1) were determined by rapid amplification of cloned ends. We used the same G-CaMP transgene for all Ca²⁺ imaging studies by crossing it into different genetic backgrounds. See Supplementary Information for details.

Ca²⁺ imaging. Ca²⁺ imaging was done on an Axiovert 200 microscope (Zeiss) under a ×40 objective. Images were acquired with a CoolSnap CCD camera (Roper) and processed by Ratiotool software (ISeeimaging). See Supplementary Information for details.

Received 7 July; accepted 19 December 2005.

- Samuel, A. D. & Sengupta, P. Sensorimotor integration: locating locomotion in neural circuits. *Curr. Biol.* **15**, R341–R343 (2005).
- Montell, C. The TRP superfamily of cation channels. *Sci STKE* **2005**, re3 (2005).

- Xu, X. Z. S. & Sternberg, P. W. A *C. elegans* sperm TRP protein required for sperm-egg interactions during fertilization. *Cell* **114**, 285–297 (2003).
- Sidi, S., Friedrich, R. W. & Nicolson, T. NompC TRP channel required for vertebrate sensory hair cell mechanotransduction. *Science* **301**, 96–99 (2003).
- Walker, R. G., Willingham, A. T. & Zuker, C. S. A *Drosophila* mechanosensory transduction channel. *Science* **287**, 2229–2234 (2000).
- Chalfie, M. & White, J. in *The Nematode Caenorhabditis elegans* (ed. Wood, W. B.) 337–391 (Cold Spring Harbor Laboratory Press, New York, 1988).
- Cronin, C. J. et al. An automated system for measuring parameters of nematode sinusoidal movement. *BMC Genet.* **6**, 5 (2005).
- Hall, W. C. & Nicolelis, M. in *Neuroscience* (eds Purves, D. et al.) 347–368 (Sinauer Associates, Sunderland, 2001).
- Goldberg, C. J. et al. in *Research Into Spinal Deformities* (ed. Stokes, I. A. F.) 53–56 (International Research Society of Spinal Deformities, Amsterdam, 1999).
- Ernst, G. G. & Chalfie, M. Genetics of sensory mechanotransduction. *Annu. Rev. Genet.* **36**, 411–453 (2002).
- Goodman, M. B. & Schwarz, E. M. Transducing touch in *Caenorhabditis elegans*. *Annu. Rev. Physiol.* **65**, 429–452 (2003).
- Sawin, E. R., Ranganathan, R. & Horvitz, H. R. *C. elegans* locomotory rate is modulated by the environment through a dopaminergic pathway and by experience through a serotonergic pathway. *Neuron* **26**, 619–631 (2000).
- Lints, R. & Emmons, S. W. Patterning of dopaminergic neurotransmitter identity among *Caenorhabditis elegans* ray sensory neurons by a TGFβ family signaling pathway and a Hox gene. *Development* **126**, 5819–5831 (1999).
- White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Phil. Trans. R. Soc. Lond. B* **314**, 1–340 (1986).
- Kahn-Kirby, A. H. et al. Specific polyunsaturated fatty acids drive TRPV-dependent sensory signaling *in vivo*. *Cell* **119**, 889–900 (2004).
- Wang, J. W., Wong, A. M., Flores, J., Vosshall, L. B. & Axel, R. Two-photon calcium imaging reveals an odor-evoked map of activity in the fly brain. *Cell* **112**, 271–282 (2003).
- Nakai, J., Ohkura, M. & Imoto, K. A high signal-to-noise Ca²⁺ probe composed of a single green fluorescent protein. *Nature Biotechnol.* **19**, 137–141 (2001).
- Way, J. C. & Chalfie, M. The *mec-3* gene of *Caenorhabditis elegans* requires its own product for maintained expression and is expressed in three neuronal cell types. *Genes Dev.* **3**, 1823–1833 (1989).
- Driscoll, M. & Chalfie, M. The *mec-4* gene is a member of a family of *Caenorhabditis elegans* genes that can mutate to induce neuronal degeneration. *Nature* **349**, 588–593 (1991).
- Richmond, J. E., Davis, W. S. & Jorgensen, E. M. UNC-13 is required for synaptic vesicle fusion in *C. elegans*. *Nature Neurosci.* **2**, 959–964 (1999).
- Cattaert, D. & Le Ray, D. Adaptive motor control in crayfish. *Prog. Neurobiol.* **63**, 199–240 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Nakai for the G-CaMP plasmid; B. Perry for help with deletion libraries; C. Cronin for assistance with data analysis; G. Schindelman, N. Moghal, C. Montell and A. Ward for comments; and G. Medina, E. Larkspur and R. Mahapatra for technical assistance. Some strains were obtained from the *Caenorhabditis* Genetics Center. This work was supported by the University of Michigan BSSP Scholar Program (X.Z.S.X.), the Helen Hay Whitney Foundation (X.Z.S.X.), and the Howard Hughes Medical Institute, of which P.W.S. is an investigator.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to X.Z.S.X. (shawnxu@umich.edu).

LETTERS

A silicon transporter in rice

Jian Feng Ma¹, Kazunori Tamai^{1*}, Naoki Yamaji^{1*}, Namiki Mitani^{1*}, Saeko Konishi², Maki Katsuhara¹, Masaji Ishiguro³, Yoshiko Murata³ & Masahiro Yano⁴

Silicon is beneficial to plant growth and helps plants to overcome abiotic and biotic stresses by preventing lodging (falling over) and increasing resistance to pests and diseases, as well as other stresses^{1–3}. Silicon is essential for high and sustainable production of rice⁴, but the molecular mechanism responsible for the uptake of silicon is unknown. Here we describe the Low silicon rice 1 (*Lsi1*) gene, which controls silicon accumulation in rice, a typical silicon-accumulating plant. This gene belongs to the aquaporin family⁵ and is constitutively expressed in the roots. *Lsi1* is localized on the plasma membrane of the distal side of both exodermis and endodermis cells, where casparian strips are located. Suppression of *Lsi1* expression resulted in reduced silicon uptake. Furthermore, expression of *Lsi1* in *Xenopus* oocytes showed transport activity for silicon only. The identification of a silicon transporter provides both an insight into the silicon uptake system in plants, and a new strategy for producing crops with high resistance to multiple stresses by genetic modification of the root's silicon uptake capacity.

Silicon is the second most abundant element in the Earth's crust and soil and is contained in significant amounts in all plants. However, plant species differ greatly in silicon accumulation, ranging from 0.1% to 10% in top dry weight^{1,6}; this difference is attributed to the difference in the ability of roots to take up silicon⁷. Rice can accumulate silicon to the level of up to 10% of shoot dry weight, which is often several times higher than that of essential macronutrients such as nitrogen, phosphate and potassium⁴. Silicon is taken up by roots in the form of silicic acid, an undissociated molecule^{8,9}. Physiological studies have shown that silicon uptake by rice roots is mediated by a type of transporter¹⁰. After it is taken up, silicon is translocated to the shoot in the form of monomeric silicic acid^{11,12} and is finally deposited on cell wall material as a polymer of hydrated, amorphous silica, forming silica–cuticle double layers and silica–cellulose double layers on the surface of leaves, stem and hulls¹³. Silicon enhances resistance of plants to diseases, pests and lodging through deposition in the apoplast and induced resistance, improves the light-interception ability by plants in a community, and minimizes transpiration losses^{1,6,14}. Therefore, plants need to accumulate large amounts of silicon.

Lsi1 (low silicon rice 1) is a rice mutant defective in silicon uptake¹⁵. This mutant accumulates less silicon in the shoot throughout its growth period compared with the wild type (Fig. 1a) and is susceptible to pests and diseases (Fig. 1b, c). The mutant has a grain yield one-tenth of that of wild-type rice (Fig. 1d). We roughly mapped the gene (*Lsi1*) controlling silicon uptake to chromosome 2 (ref. 16). For fine mapping of *Lsi1*, we used about 1,000 homozygotes with low silicon uptake, which were selected from F₂ plants derived from a cross between *Lsi1* and an indica cultivar Kasalath. We developed new markers and mapped the candidate region of *Lsi1* to 88 kilobases (kb) between the markers *Lsi1*-4 and E60168 and to

13.9 kb between the markers *Lsi1*-a and *Lsi1*-6 (Fig. 2a, b). Using gene prediction software we predicted a gene in the candidate region (Fig. 2b), sequenced it and made comparisons between the wild type and *Lsi1* mutant. We found a mutation in the DNA sequence of the candidate gene (G in the wild type; A in the mutant) that results in an amino acid change from alanine in the wild type to threonine in the mutant at position 132 (Fig. 2d). Thus, we considered this candidate gene to be *Lsi1*. The gene consists of five exons and four introns (Fig. 2c). The complementary DNA of this gene is 1,409-base-pairs (bp) long and the deduced protein comprises 298 amino acids (Fig. 2d).

The gene is predicted to encode a membrane protein similar to water channel proteins (aquaporins)⁵. The predicted amino acid sequence has six transmembrane domains and two Asn-Pro-Ala (NPA) motifs, which is well conserved in typical aquaporins (Fig. 2d). BLAST search and ClustalW analysis revealed that *Lsi1* belongs to a Nod26-like major intrinsic protein (NIP) subfamily (Fig. 2f). We found three close homologues in maize (ZmNIP2-1,

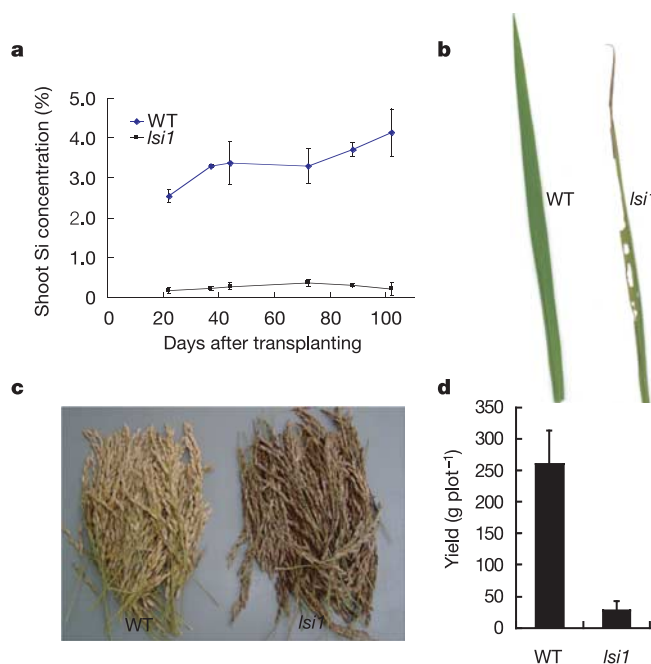


Figure 1 | Phenotype of the *Lsi1* mutant. **a**, Silicon concentration of shoots at each growth stage in wild-type rice (WT; cv. Oochikara) and an *Lsi1* mutant grown in a field. **b**, A mature leaf showing pest damage in the mutant due to low silicon. **c**, Panicles at harvest showing that panicles of the *Lsi1* mutant are infected by diseases, resulting in discoloration. **d**, Rice grain yield per plot (70 cm × 70 cm). Data are means ± s.d. (*n* = 3).

¹Research Institute for Bioresources, Okayama University, Chuo 2-20-1, Kurashiki 710-0046, Japan. ²Institute of Society for Techno-innovation of Agriculture, Forestry and Fisheries, Kamiyokoba, Tsukuba, Ibaraki 305-0854, Japan. ³Suntory Institute for Bioorganic Research, 1-1-1 Wakayamadai, Shimamoto-cho, Mishima-gun, Osaka 618-8503, Japan. ⁴Molecular Genetics Department, National Institute of Agrobiological Sciences, Tsukuba, Ibaraki 305-8602, Japan.

*These authors contributed equally to this work.

ZmNIP2-2 and ZmNIP2-3) sharing 77–83% sequence identity¹⁷, and one homologue in rice (Os06g12310, named *Lsi6*) with 77% identity (Fig. 2f). However, the transport selectivity and physiological roles of ZmNIP2—as with all members of the subgroup—are unknown. We show here that one of the NIPs, *Lsi1*, is a transporter for silicic acid in rice roots (Supplementary Fig. 1). Maize is also able to accumulate silicon, suggesting that ZmNIP2-1, ZmNIP2-2 and ZmNIP2-3 might be involved in silicon uptake. Alanine at position 132 seems to be a critical residue, because substitution of this amino acid in the mutant significantly alters the conformation according to the modelling of the native and mutant proteins. Thus, substitution of Ala for Thr at position 132 (Ala132Thr) provoked severe steric interactions with Val 55 and Val 59 in helix 1 (H1), facilitating a movement of H1. This unfavourable interaction would affect the conformation of Asn 108, the pore-forming residue in the P-loop (Fig. 2e).

We found that *Lsi1* was mainly expressed in the roots (Fig. 3a).

This expression was constitutive, but regulated by silicon level; the expression was decreased by one-quarter by continuous silicon supply for 3 days (Fig. 3b). Results of *in situ* hybridization showed that *Lsi1* messenger RNA was localized at the exodermis and endodermis (Supplementary Fig. 2). We investigated the subcellular localization of *Lsi1* by delivering a translational fusion between *Lsi1* and green fluorescent protein (GFP) into onion epidermal cells by particle bombardment. Cells expressing the *Lsi1*–GFP fusion showed a GFP signal only at the plasma membrane (Supplementary Fig. 3), whereas the signal for cells expressing GFP alone was found in the nucleus and cytosol.

To examine further the localization of *Lsi1* in rice roots, we generated transgenic rice plants carrying the open reading frame for *Lsi1* fused with GFP under the control of the *Lsi1* promoter region (2 kb). The GFP fluorescence signal was observed in the main and lateral roots, but not in root hairs (Fig. 3c, d). This is consistent with

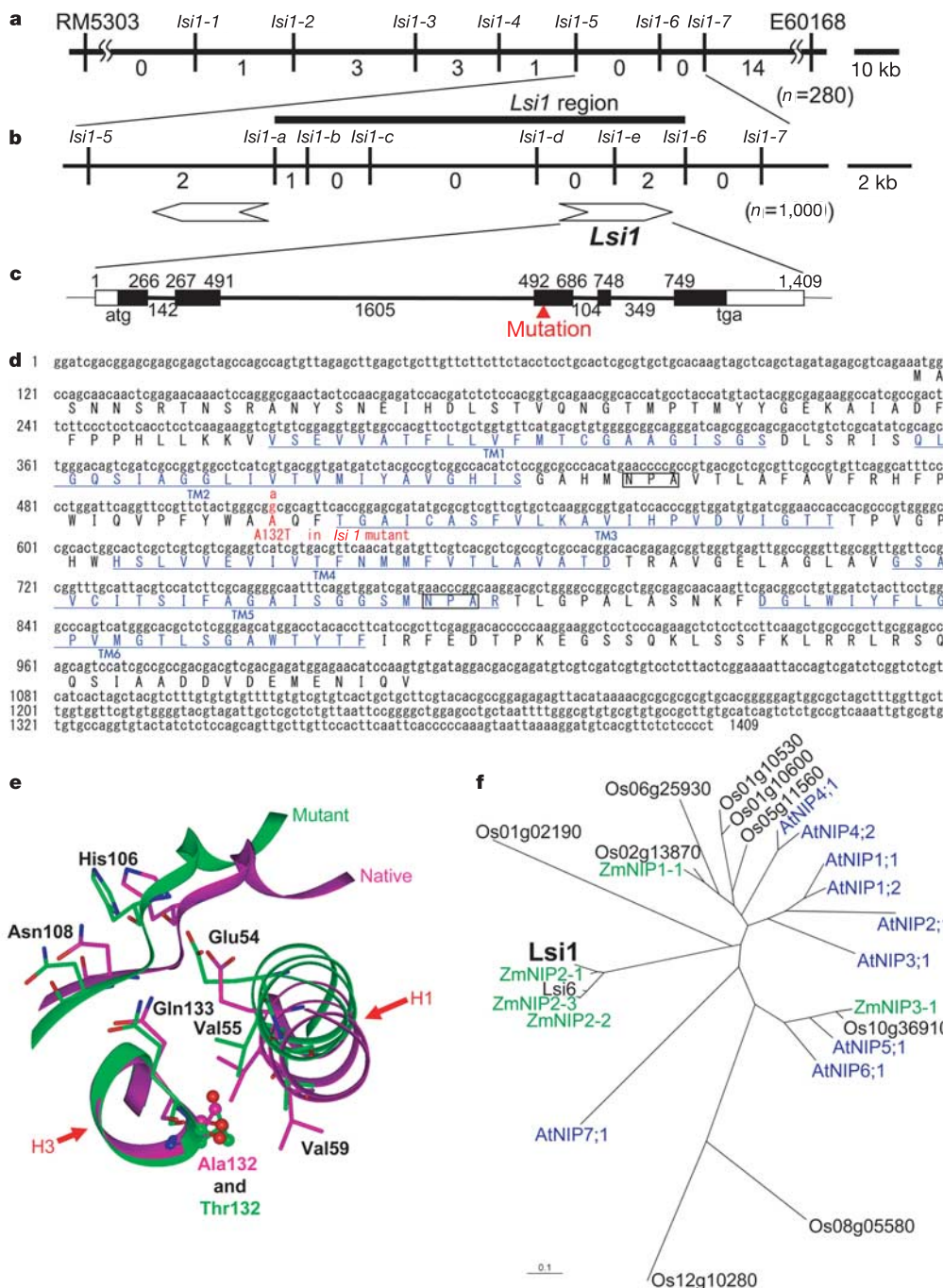
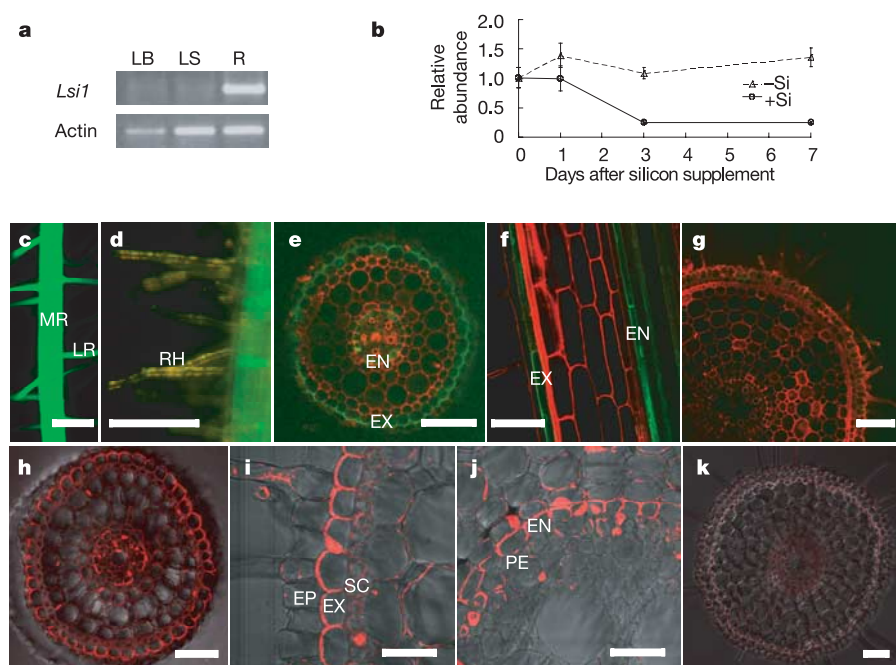


Figure 2 | Mapping of *Lsi1* and gene structure. **a, b**, The *Lsi1* mutation was mapped on the long arm of chromosome 2 between markers RM5303 and E60168. The number of recombinants between the molecular markers is indicated below the high-resolution map. **c**, *Lsi1* gene structure at genomic sequence. Five exons are boxed; a black box shows the open reading frame. **d**, *Lsi1* cDNA and predicted amino acid sequence based on full-length cDNA clone (accession number AK069842). Red letters show position of *Lsi1* mutation. Predicted transmembrane domains (TM) are underlined and the conserved NPA motif is boxed. **e**, Model of the *Lsi1* protein from the wild-type and mutant rice. **f**, Phylogenetic relationship of *Lsi1* proteins in rice (black), *Arabidopsis* (blue) and maize (green).

**Figure 3 | Expression and localization of *Lsi1*.**

a, Expression analysis of *Lsi1* and Actin (internal standard) in the leaf blade (LB), leaf sheath (LS) and root (R). **b**, Time-dependent expression of *Lsi1*. Data are means $\pm$ s.d. ($n = 3$).

c–f, Fluorescence of the *Lsi1*–GFP fusion protein in transgenic plants. The main root (MR), lateral root (LR), root hair (RH), endodermis (EN) and exodermis (EX) are shown. **c, d**, Stereoscopic microscope images. Scale bars, 500 μ m. **e, f**, Cross-section (**e**) and longitudinal section (**f**) observed by laser scanning confocal microscopy, counterstained with propidium iodide. **g**, Control (wild-type root). Scale bar, 50 μ m. **h–j**, *Lsi1* immunolocalization stained with anti-*Lsi1* polyclonal antibody. Epidermis (EP), exodermis, sclerenchyma (SC), endodermis and pericycle (PE) are shown. **k**, Control (non-immunized serum). Scale bars, 50 μ m (**h, k**) and 20 μ m (**i, j**).

the results of a previous physiological study that root hairs do not have any demonstrable role in silicon uptake, but that lateral roots contribute significantly to silicon uptake¹⁸. Within a root, we found that the GFP fluorescence signal was observed on the plasma membrane of both exodermis and endodermis (Fig. 3e, f), where casparian strips exist (Supplementary Fig. 4). To confirm this sub-cellular localization of *Lsi1*, we stained the roots with an anti-*Lsi1* polyclonal antibody. *Lsi1* was localized on the plasma membrane of the distal side of both exodermis and endodermis cells, which is similar to the localization seen for the transgenic plants carrying the GFP fusion (Fig. 3h–j). Because solutes are unable to pass casparian strips freely¹⁹, transporters are needed to reach the stele for translocation from the roots to the shoot. Localization of *Lsi1* at the plasma membrane of the distal side of both exodermis and endodermis cells indicates that *Lsi1* is a transporter responsible for silicon uptake, and therefore for silicon accumulation in rice.

Because plants were not generated from calluses of the cultivar (Oochikara) used, probably due to the specificity of this cultivar, we could not produce transgenic plants by introducing *Lsi1* into the silicon-uptake-defective (*lsi1*) mutant. Therefore, we used RNA interference (RNAi) to suppress the expression of *Lsi1* in the Nipponbare cultivar, which has the same gene and same silicon uptake capacity as the Oochikara cultivar. In RNAi transgenic lines, silicon uptake was significantly reduced compared with vector control plants; expression of *Lsi1* in RNAi transgenic lines was significantly suppressed (Fig. 4a, b). The RNAi transgenic lines also showed higher resistance to Ge toxicity (Supplementary Fig. 5); however, there was no difference in water uptake between RNAi lines and vector control lines (data not shown).

We also investigated the silicon transport activity of *Lsi1* by injecting cRNA encoding *Lsi1* or injecting water into *Xenopus laevis* oocytes. We measured the silicon or glycerol inside the oocytes after incubation in a solution containing silicic acid or glycerol for 30 min. Oocytes expressing *Lsi1* had a silicon uptake rate 2.4 times greater than the control (water-injected) oocytes (Fig. 4c). However, the difference in the transport activity for glycerol was very small between control oocytes and those injected with cRNA encoding *Lsi1*. Furthermore, the transport activity of silicon was unaffected by the presence of equimolar amounts of glycerol (data not shown), suggesting that *Lsi1* is specific for silicon transport. Kinetic analysis showed that silicon transport activity increased with increasing silicon concentrations in external solution (Fig. 4d).

Our results support the concept that *Lsi1* is a transporter for silicon in rice roots. Silicon is ubiquitous in the environment, and all living organisms take up silicon. However, the genes responsible for silicon uptake have not been identified so far in higher plants. A gene family encoding silicon transporters has been identified in the marine diatom *Cylindrotheca fusiformis*, which requires silicon as an essential element^{20,21}; however, this gene has no similarity to *Lsi1*. Furthermore, silicon uptake was not increased by introducing one of the diatom silicon transporter genes into tobacco²², indicating that the silicon uptake system in higher plants is different compared with that in diatoms. To our knowledge, this is the first report of a silicon transporter in higher plants, and it provides a molecular basis for their silicon transport system. Many plants, especially dicotyledonous species, are unable to accumulate silicon in sufficient amounts to be beneficial. Identification of *Lsi1* may provide a new strategy for producing plants with resistance to multiple stresses by genetic

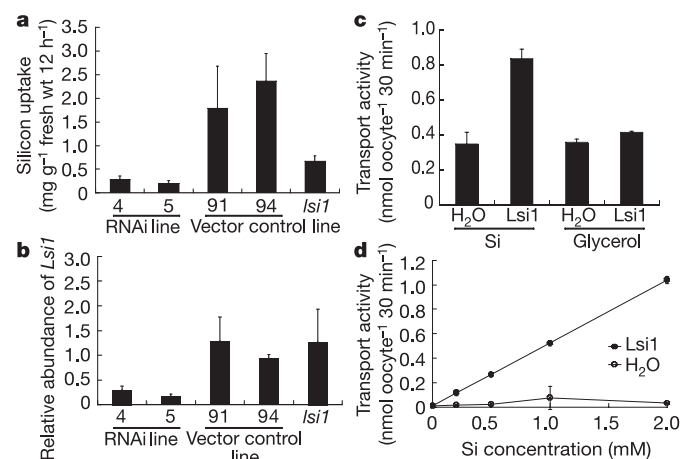


Figure 4 | Transport activity of *Lsi1*. **a**, Effect of *Lsi1* suppression by RNAi silencing on silicon uptake. **b**, Relative root transcript level of *Lsi1*. Actin was used as an internal control and relative value to vector control is shown. **c**, Uptake of silicon and glycerol by *Xenopus* oocytes that were injected with water or *Lsi1* cRNA. **d**, Kinetics of silicon uptake by *Xenopus* oocytes injected with water or *Lsi1* cRNA. Oocytes were exposed to a solution containing different silicon concentrations labelled with ⁶⁸Ge for 30 min. Data are means $\pm$ s.d. ($n = 3$).

modification of the root's silicon uptake capacity. As silicon has been implicated in optimal bone and connective tissue development in the human body²³, enhanced silicon uptake in plants may also result in increased silicon accumulation in food, thereby improving silicon nutrition in humans.

METHODS

Plant growth. We grew wild-type (cv. Oochikara) and mutant (*lsi1*) rice in a field from May to September in 2003. We obtained shoot samples at various growth stages and determined the silicon accumulation as described previously²⁴. Grain yield per plot (70 cm × 70 cm) was investigated at harvest.

Map-based cloning of *Lsi1*. We evaluated about 4,000 F₂ plants derived from a cross between *lsi1* and the Kasalath cultivar for silicon uptake, and selected about 1,000 homozygote plants showing low silicon uptake for high-resolution mapping. Linkage analysis using both SSR and PCR markers revealed that *Lsi1* was restricted to a 13.9-kb genomic region on the bacterial artificial chromosome (BAC) clone OJ1118_G04. Only one gene was predicted in this region by rice GAAS, and the sequence of this gene was determined for both the wild type and the mutant.

Immunohistological fluorescence staining. Rice roots were fixed in 4% (w/v) paraformaldehyde and 60 mM sucrose and then embedded in 5% agar. Sections sliced to 50-μm thickness were incubated in PBS containing 0.1% (w/v) pectolyase Y-23 and then in PBS containing 0.3% (v/v) Triton X-100. The nonspecific reaction was blocked with 5% (w/v) BSA in PBS. Then we incubated the slides with purified rabbit anti-*Lsi1* polyclonal antibodies and subsequently with secondary antibodies (Alexa Fluor 555 goat anti-rabbit IgG; Molecular Probes). We observed the sections with a laser scanning confocal microscope (LSM510; Zeiss).

Generation of transgenic rice. To investigate the cellular and subcellular localization of *Lsi1*, we introduced a construct consisting of the promoter (2 kb) and *Lsi1* cDNA fused with GFP to calluses (cv. Nipponbare) using an *Agrobacterium*-mediated transformation system²⁵. We selected transformed calluses by hygromycin resistance, and from them regenerated plants. We examined fluorescence in the transgenic rice roots by laser-scanning confocal microscopy after counterstaining with propidium iodide.

To generate the hairpin RNAi construct, we cloned two copies of a 299-bp fragment (15–313 bases from transcriptional start) of *Lsi1* cDNA at inverted repeats into the pHELLSGATE vector under control of a 35S promoter²⁶ and subsequently introduced it to calluses (cv. Nipponbare) as described above. We measured the silicon uptake in two independent RNAi transgenic lines as well as in two independent vector control lines and the mutant *lsi1* as described¹⁵, and examined the expression level of *Lsi1* with real-time PCR as described below.

Real-time PCR. We extracted total RNA from the leaf blade, leaf sheath and root supplied with or without silicon for 1, 3 and 7 days or from transgenic rice roots and then converted it to cDNA. The *Lsi1* and Actin (internal control) cDNAs were amplified using SYBR green I real-time PCR with pairs of primers: *Lsi1*, 5'-CGGTGGATGTGATCGGAACCA-3' (forward) and 5'-CGTCGAACCTT GTTGCTCGCCA-3' (reverse); Actin, 5'-GACTCTGGTGATGGTGCAGC-3' (forward) and 5'-GGCTGGAAGAGGACCTCAGG-3' (reverse).

Transport activity in oocytes. We performed *in vitro* transcription from pX8G-ev1 poly(A)⁺ vector carrying *Lsi1* cDNA. We injected the capped cRNA (50 nl, 1 ng nl⁻¹) or distilled water into *Xenopus* oocytes selected according to size and developmental stage. After a 1-day incubation, the oocytes were exposed to a solution containing 2 mM silicon as silicic acid or 2 mM glycerol labelled with ¹⁴C (40 MBq mmol⁻¹) for 30 min. The oocytes were then washed with a buffer without silicon and homogenized with 0.1 N HNO₃. We determined the concentration of silicon in the lysates as described previously²⁴ and that of glycerol by a liquid scintillation counter. The kinetics of silicon uptake was performed by exposing the oocytes to a solution containing different silicon concentrations ranging from 0.2 to 2.0 mM labelled with ⁶⁸Ge (10 MBq mmol⁻¹) following the method reported²⁷. After 30 min, we washed the oocytes with the solution without silicon five times and then measured the radioactivity by a liquid scintillation counter 24 h later.

Modelling of the *Lsi1* structure. We used the crystal structure of the monomeric structure of aquaporin (Protein Data Bank 1J4N) to build the model of the monomer of *Lsi1*. Sequence alignment and the initial homology modelling were performed with the homology module installed in Insight II (version 2000, Molecular Simulations Inc.). Si(OH)₄ was roughly docked into the substrate-filtering site. We minimized the initial complex model and optimized the whole structure of the complex model by the molecular dynamics/minimization procedure without any structural constraints. We selected the lowest-energy structure as an energy-refined complex model. We generated the Ala132Thr mutant by replacing Ala with Thr and optimized the mutant structure.

Received 5 September 2005; accepted 18 January 2006.

- Epstein, E. The anomaly of silicon in plant biology. *Proc. Natl Acad. Sci. USA* **91**, 11–17 (1994).
- Ma, J. F. Role of silicon in enhancing the resistance of plants to biotic and abiotic stresses. *Soil Sci. Plant Nutr.* **50**, 11–18 (2004).
- Richmond, K. E. & Sussman, M. Got silicon? The non-essential beneficial plant nutrient. *Curr. Opin. Plant Biol.* **6**, 268–272 (2003).
- Savant, N. K., Snyder, G. H. & Datnoff, L. E. Silicon management and sustainable rice production. *Advan. Agron.* **58**, 151–199 (1997).
- Luu, D. T. & Maurel, C. Aquaporins in a challenging environment: molecular gears for adjusting plant water status. *Plant Cell Environ.* **28**, 85–96 (2005).
- Ma, J. F. & Takahashi, E. in *Soil, Fertilizer, and Plant Silicon Research in Japan* (Elsevier, Amsterdam, 2002).
- Takahashi, E., Ma, J. F. & Miyake, Y. The possibility of silicon as an essential element for higher plants. *Comments Agric. Food Chem.* **2**, 99–122 (1990).
- Takahashi, E. & Hino, K. Silicon uptake by plants with special reference to the forms of dissolved silicon. *J. Sci. Soil Manure Jpn* **49**, 357–360 (1978).
- Raven, J. A. in *Silicon in Agriculture* (eds Datnoff, L. E., Snyder, G. H. & Korndörfer, G. H.) 41–55 (Elsevier, Amsterdam, 2001).
- Tamai, K. & Ma, J. F. Characterization of silicon uptake by rice roots. *New Phytol.* **158**, 431–436 (2003).
- Casey, W. H., Kinrade, S. D., Knight, C. T. G., Rains, D. W. & Epstein, E. Aqueous silicate complexes in wheat, *Triticum aestivum* L. *Plant Cell Environ.* **27**, 51–54 (2003).
- Mitani, N., Ma, J. F. & Iwashita, T. Identification of silicon form in the xylem of rice (*Oryza sativa* L.). *Plant Cell Physiol.* **46**, 279–283 (2005).
- Yoshida, S. Chemical aspects of the role of silicon in physiology of the rice plant. *Bull. Natl Inst. Agric. Sci. B* **15**, 1–58 (1965).
- Fauteux, F., Remus-Borel, W., Menzies, J. G. & Belanger, R. R. Silicon and plant disease resistance against pathogenic fungi. *FEMS Microbiol. Lett.* **249**, 1–6 (2005).
- Ma, J. F., Tamai, K., Ichii, M. & Wu, K. A rice mutant defective in Si uptake. *Plant Physiol.* **130**, 2111–2117 (2002).
- Ma, J. F. *et al.* Characterization of Si uptake system and molecular mapping of Si transporter gene in rice. *Plant Physiol.* **136**, 3284–3289 (2004).
- Chamont, F., Barrieu, F., Wojcik, E., Chrispeels, M. J. & Jung, R. Aquaporins constitute a large and highly divergent protein family in maize. *Plant Physiol.* **125**, 1206–1215 (2001).
- Ma, J. F., Goto, S., Tamai, K. & Ichii, M. Role of root hairs and lateral roots in silicon uptake by rice. *Plant Physiol.* **127**, 1773–1780 (2001).
- Taiz, L. & Zeiger, E. *Plant Physiology* 103–124 (Sinauer, Sunderland, 1998).
- Hildebrand, M., Higgins, D. R., Busser, K. & Volcani, B. E. Silicon-responsive cDNA clones isolated from the marine diatom *Cylindrotheca fusiformis*. *Gene* **132**, 213–218 (1993).
- Hildebrand, M., Volcani, B. E., Gassmann, W. & Schroeder, J. I. A gene family of silicon transporters. *Nature* **385**, 688–689 (1997).
- Ma, J. F. Mechanism of Si uptake in plants. *Fertilizer* **94**, 26–32 (2003).
- Sripanyakorn, S., Jugdaohsingh, R., Thompson, R. P. H. & Powell, J. J. Dietary silicon and bone health. *Nutr. Bull.* **30**, 222–230 (2005).
- Ma, J. F., Higashitani, A., Sato, K. & Tateda, K. Genotypic variation in Si content of barley grain. *Plant Soil* **249**, 383–387 (2003).
- Hiei, Y., Ohta, S., Komari, T. & Kumashiro, T. Efficient transformation of rice (*Oryza sativa* L.) mediated by *Agrobacterium* and sequence analysis of the boundaries of the T-DNA. *Plant J.* **6**, 271–282 (1994).
- Helliwell, C. A., Wesley, S. V., Wielopolska, A. J. & Waterhouse, P. M. High-throughput vectors for efficient gene silencing in plants. *Funct. Plant Biol.* **29**, 1217–1225 (2002).
- Tallberg, P., Koski-Vahala, J. & Hartikainen, H. Germanium-68 as a tracer for silicon fluxes in freshwater sediment. *Water Res.* **36**, 956–962 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This research was supported by a Grant-in-Aid for Scientific Research on Priority Areas from the Ministry of Education, Culture, Sports, Science and Technology of Japan (to J.F.M.) and a grant from the Ministry of Agriculture, Forestry and Fisheries of Japan (Rice Genome Project, to J.F.M.).

Author Contributions K.T., N.Y. and N.M. contributed equally to this work. K.T. cloned the gene *Lsi1*, N.Y. investigated the localization of *Lsi1*, and N.M. measured the transport activity of *Lsi1*. J.F.M. performed the field and RNAi experiments and wrote the paper. All authors discussed the results and commented on the manuscript.

Author Information The nucleotide sequence data reported in this paper has been deposited in the DDBJ/EMBL/GenBank nucleotide sequence databases under accession number AB222272. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.F.M. (maj@rib.okayama-u.ac.jp).

LETTERS

Regulation of cancer cell migration and bone metastasis by RANKL

D. Holstead Jones^{1,2,3,*†}, Tomoki Nakashima^{1*}, Otto H. Sanchez^{4†}, Ivona Kozieradzki^{1,2,3}, Svetlana V. Komarova⁵, Ildiko Sarosi⁶, Sean Morony⁶, Evelyn Rubin^{2,3}, Renu Sarao¹, Carlo V. Hojilla⁴, Vukoslav Komnenovic¹, Young-Yun Kong⁷, Martin Schreiber⁸, S. Jeffrey Dixon⁹, Stephen M. Sims⁹, Rama Khokha^{2,4}, Teiji Wada¹ & Josef M. Penninger^{1,2,3}

Bone metastases are a frequent complication of many cancers that result in severe disease burden and pain^{1–3}. Since the late nineteenth century, it has been thought that the microenvironment of the local host tissue actively participates in the propensity of certain cancers to metastasize to specific organs, and that bone provides an especially fertile ‘soil’⁴. In the case of breast cancers, the local chemokine milieu is now emerging as an explanation for why these tumours preferentially metastasize to certain organs⁵. However, as the inhibition of chemokine receptors *in vivo* only partially blocks metastatic behaviour⁶, other factors must exist that regulate the preferential metastasis of breast cancer cells. Here we show that the cytokine RANKL (receptor activator of NF- κ B ligand)^{7,8} triggers migration of human epithelial cancer cells and melanoma cells that express the receptor RANK. RANK is expressed on cancer cell lines and breast cancer cells in patients. In a mouse model of melanoma metastasis⁹, *in vivo* neutralization of RANKL by osteoprotegerin results in complete protection from paralysis and a marked reduction in tumour burden in bones but not in other organs. Our data show that local differentiation factors such as RANKL have an important role in cell migration and the tissue-specific metastatic behaviour of cancer cells.

RANKL (also referred to as OPG, TRANCE or ODF) is a member of the tumour necrosis factor (TNF) family of cytokines that binds to its receptor RANK to control osteoclast differentiation, activation and survival^{7,8,10}. Osteoprotegerin (OPG) is a soluble decoy receptor for RANKL that blocks ligand binding to RANK, thereby preventing the signalling required for osteoclast differentiation and activation¹¹. RANK is also constitutively expressed in normal mammary gland epithelial cells, but RANKL expression is induced by sex hormones during pregnancy¹². Genetically, both RANKL and RANK are essential for the development of the lactating mammary gland during pregnancy¹² and for lymph node organogenesis in mouse embryos⁷.

This unexpected distribution of RANKL led us to examine multiple epithelial tissues for RANK expression. In all mouse epithelial tissues analysed, RANK messenger RNA was present, even in epithelial tissues of the early embryo. We also detected marked RANK expression in a large number of primary human breast tumour samples as well as in cancer cells present in local lymph node metastases (Fig. 1a and Supplementary Fig. 1a). Moreover, several human prostate and breast cancer cell lines, but not colon cancer cell lines, expressed RANK mRNA (Supplementary Fig. 1b),

and RANK protein was detected on the surface of breast cancer cells by fluorescein isothiocyanate (FITC)-labelled RANKL binding (Fig. 1b and Supplementary Fig. 2a). Stimulation of RANK-positive human breast cancer cells with recombinant RANKL induced strong actin polymerization that could be blocked by OPG (Fig. 1c). Although RANK stimulation resulted in enhanced activation of protein kinase B (PKB/AKT) and extracellular signal-regulated kinases 1 and 2 (ERK1/2) (Supplementary Fig. 2b, c), RANKL had no apparent effect on proliferation or death susceptibility of these epithelial tumour cells (Supplementary Fig. 2d, e). Thus, RANK is expressed on many different epithelial tissues and epithelial tumour cells, and can activate specific downstream signalling pathways.

As actin polymerization is a hallmark of chemokine receptor signalling in cancer cell lines^{6,13}, we speculated that RANKL and RANK might have a role in epithelial cell migration. *In vitro* stimulation of three different human breast cancer cell lines (MDA-MB-231, MCF-7 and Hs578T) with RANKL resulted in concentration-dependent cell migration, which was blocked using the decoy receptor OPG (Fig. 1d). Furthermore, RANKL triggered migration of two RANK-expressing prostate cancer cell lines, but had a negligible effect on migration of the colon cancer cell line Colo205, in which we failed to detect RANK expression (Fig. 1e). The extent of RANKL-induced migration in the breast cancer cell line MDA-MB-231 was comparable to the previously reported migration of these cells in response to the chemokines 6Ckine and CTAK⁶, but less than that observed with stromal cell-derived factor 1 α (SDF-1 α) (Fig. 1f and Supplementary Fig. 3). Although OPG inhibited RANKL-induced migration, it had no apparent effect on chemokine-induced migration (Fig. 1f). Thus, RANKL induces migration of malignant epithelial cells expressing RANK.

As physiological and malignant cell invasion use similar molecular mechanisms¹⁴, we evaluated the effects of RANKL on the migration of primary, non-transformed cells that express RANK. RANKL-induced cell migration was observed in primary mouse mammary epithelial cells freshly isolated from virgin females, and the primary, non-transformed mammary epithelial cell line MCF10A (Fig. 2a). Moreover, RANKL triggered directional migration of mature osteoclasts towards a RANKL source (see Supplementary Information and Supplementary Fig. 4). These data identify a role for the TNF/TNFR family molecules RANKL and RANK in the migration of primary breast epithelial cells and epithelial tumour cells.

¹IMBA, Institute of Molecular Biotechnology of the Austrian Academy of Sciences, Dr. Bohr Gasse 3, A-1030 Vienna, Austria. ²Department of Medical Biophysics and ³Department of Immunology, University of Toronto, 610 University Avenue, Toronto, Ontario M5G 2C1, Canada. ⁴Ontario Cancer Institute, University Health Network, 610 University Avenue, Toronto, Ontario M5G 2C1, Canada. ⁵Faculty of Dentistry, McGill University, Montreal, Quebec H3A 1A4, Canada. ⁶Amgen Inc., Thousand Oaks, California 91320-1799, USA. ⁷Division of Molecular and Life Sciences, Pohang University of Science and Technology, Pohang, 790-784 Kyungbuk, South Korea. ⁸Medical University of Vienna, Department of Obstetrics and Gynecology, Waehringer Guertel 18-20, A-1090 Vienna, Austria. ⁹CIHR Group in Skeletal Development and Remodeling, Department of Physiology and Pharmacology and Division of Oral Biology, Schulich School of Medicine & Dentistry, The University of Western Ontario, London, Ontario N6A 5C1, Canada. [†]Present address: Faculty of Health Sciences, University of Ontario Institute of Technology, 2000 Simcoe Street North, Oshawa, Ontario L1H 7K4, Canada.

*These authors contributed equally to this work.

In contrast to most metastatic breast cancers², some tumours, including malignant melanomas, metastasize to bone without stimulating osteoclastic resorption⁹. We found that a mouse B16F10 melanoma subclone¹⁵ expresses high levels of RANK mRNA (Supplementary Fig. 5a) and protein, as determined by RANKL-FITC binding (Fig. 2b). Similar to our results in breast and prostate cancer cells, RANKL had no apparent effects on proliferation or cell death in response to anisomycin, sorbitol or irradiation with ultraviolet light (not shown). Stimulation of B16F10 cells with RANKL caused actin polymerization (Fig. 2c) and increased cell migration (Fig. 2d). RANKL-induced migration of B16F10 melanoma cells was dependent on the concentration of

RANKL and could be inhibited by the decoy receptor OPG (Fig. 2d). Stimulation of B16F10 cells with colony-stimulating factor-1 (CSF-1), which is required for RANKL-mediated osteoclastogenesis⁸, had no apparent effect on the migration of melanoma cells (Fig. 2d). The extent of RANKL-induced migration in B16F10 melanoma cells was comparable to the migration of these cells in response to the chemokines 6Ckine, CTACK and SDF-1 α (Supplementary Fig. 5b). Moreover, the effects of RANKL were additive with 6Ckine and CTACK, but not with SDF-1 α (Supplementary Fig. 5b), suggesting potential synergies between RANK and chemokine signalling that warrant further investigation. Treatment with OPG inhibited RANKL-, but not SDF-1 α -induced migration,

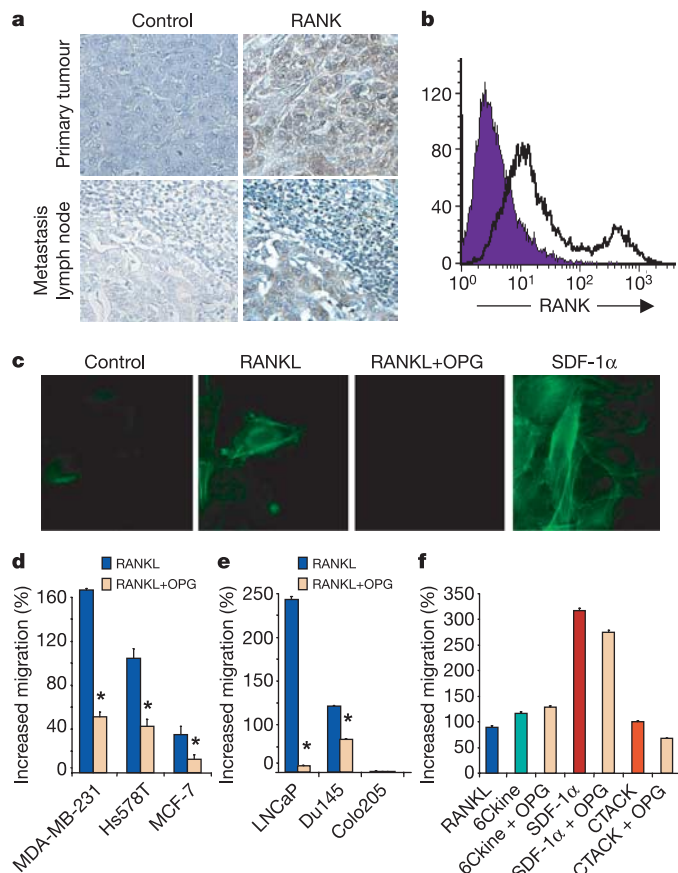


Figure 1 | RANK is expressed on breast cancer cells in patients and mediates migration of epithelial tumour cells. **a**, Expression of RANK on breast cancer cells at the site of the primary tumour and in lymph node metastases. Human breast and lymph node tissue arrays were stained with anti-RANK or control antibodies. Representative data are shown. Original magnification $\times 20$. **b**, Expression of RANK on MDA-MB-231 breast cancer cells. Background staining is shown in purple. **c**, rRANKL (2.5 μ g ml⁻¹) and SDF-1 α (80 ng ml⁻¹) trigger actin polymerization (detected by phalloidin-FITC) in MDA-MB-231 cells. OPG (10 μ g ml⁻¹) blocks RANKL-induced actin polymerization. **d**, Migration of MDA-MB-231, Hs578T and MCF-7 human breast cancer cells in response to rRANKL (2.5 μ g ml⁻¹); $n = 10$ experiments. Asterisk, OPG (10 μ g ml⁻¹) significantly reduced migration in MDA-MB-231 ($P < 0.001$), Hs578T ($P < 0.0001$) and MCF-7 ($P < 0.005$) cells. **e**, Migration of LNCaP and Du145 human prostate carcinoma cells and Colo205 human colon cancer cells in response to rRANKL (2.5 μ g ml⁻¹); $n = 3$ experiments. Asterisk, OPG (10 μ g ml⁻¹) significantly reduced migration of LNCaP ($P < 0.02$) and Du145 ($P < 0.05$) cells. **f**, Migration of MDA-MB-231 cells in response to rRANKL (2.5 μ g ml⁻¹) and the chemokines 6Ckine (120 ng ml⁻¹), SDF-1 α (80 ng ml⁻¹) and CTACK (100 ng ml⁻¹). There was no significant effect of rOPG (10 μ g ml⁻¹) on chemokine-induced cell migration ($n = 3$ experiments). In **d–f**, the percentage increased migration ($\pm$ s.d.) compared to unstimulated control cells is shown.

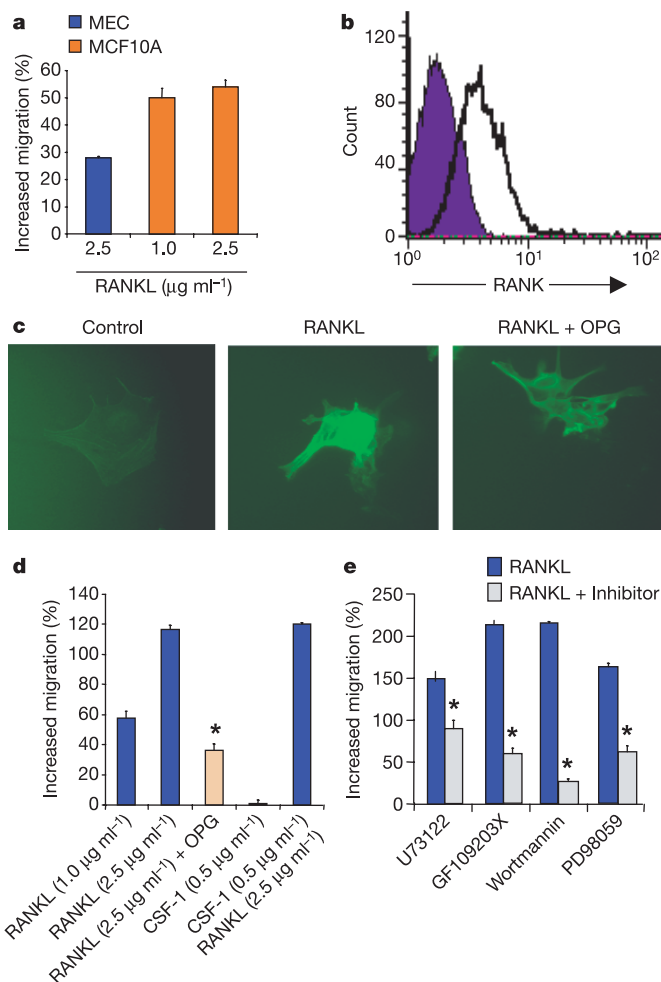


Figure 2 | RANKL triggers migration of normal mammary epithelial cells and murine B16F10 melanoma cells. **a**, Migration of freshly isolated mouse mammary gland epithelial cells (MEC) and non-transformed human MCF10A breast epithelial cells in response to rRANKL. Percentage increased migration ($\pm$ s.d.) compared to non-stimulated control cells is shown. **b**, Surface expression of RANK on mouse B16F10 melanoma cells. Background staining is shown in purple. **c**, rRANKL (2.5 μ g ml⁻¹) triggers actin polymerization (phalloidin-FITC) in B16F10 cells that is blocked by rOPG (10 μ g ml⁻¹). **d**, Migration ($\pm$ s.d.) of B16F10 cells in response to rRANKL in the absence or presence of rOPG (10 μ g ml⁻¹). Migration in response to CSF-1 is shown as a control. One result representative of ten experiments using different stimulation conditions is shown. Asterisk, $P < 0.0004$ between samples treated with rRANKL and rRANKL + OPG. **e**, Migration ($\pm$ s.d.) of B16F10 cells in response to RANKL (2.5 μ g ml⁻¹) in the presence or absence of the inhibitors U73122 (10 μ M, PLC blocker), GF109203X (10 μ M, PKC blocker), wortmannin (100 nM, PI(3)K blocker) or PD98059 (10 μ M, MEK1/2 blocker). Asterisk, $P < 0.05$ between samples treated with rRANKL and rRANKL + inhibitor.

whereas inhibition of the SDF-1 α receptor with anti-CXCR4 antibody blocked SDF-1 α -, but not RANKL-induced cell migration (Supplementary Fig. 5c). Phospholipase C (PLC), protein kinase C (PKC), ERK and phosphatidylinositol-3-OH kinase (PI(3)K) pathways have all been shown to be essential for chemokine-receptor induced cell migration, and we have recently shown that RANKL signals through PLC in osteoclasts¹⁶. Both RANKL and SDF-1 α stimulation of B16F10 melanoma cells induced actin polymerization (Fig. 2c) and ERK1/2 phosphorylation (Supplementary Fig. 5d), indicating that RANK and chemokine receptors use similar downstream signalling pathways. Inhibition of these signalling pathways inhibited RANKL-induced migration of this B16F10 melanoma subclone (Fig. 2e). These results indicate that in addition to chemokines, RANKL regulates migration of the mouse melanoma cell line B16F10.

To determine whether RANKL/RANK-regulated migration of cancer cells has a role in tumour metastasis *in vivo*, we analysed whether inhibition of RANKL/RANK through the decoy receptor OPG altered the metastasis into the bones. Intracardiac injection of mouse B16F10 melanoma cells into the left cardiac ventricle has previously been established as an *in vivo* model system to study metastasis into several organs, including the adrenal glands, the choroid plexus of the brain, the ovaries and bone⁹. Notably, the B16F10 subclone used in our experiments does not trigger osteoclast activation, a feature that allowed us to uncouple the direct effects of

RANKL on tumour metastasis from osteoclast-mediated effects¹⁷ (Supplementary Fig. 6a–c). Moreover, the tumour burden of B16F10 cells in vertebrae correlates with spinal invasion and paralysis as a functional disease read-out of bone metastasis (Fig. 3i and Supplementary Fig. 7a–c).

Injection of B16F10 melanoma cells into syngeneic C57BL/6 mice resulted in rapid metastasis of melanin-producing cancer cells into all long bones (Fig. 3a–c), vertebrae (Supplementary Fig. 7b), ovaries (Supplementary Fig. 7e), adrenal glands (Supplementary Fig. 7h) and the choroid plexus of the brain (not shown). Metastases were observed macroscopically in all animals analysed at days 12, 14 and 17 after injection, and virtually all bones in B16F10-injected mice showed black colour owing to the melanin-producing tumour cells (Fig. 3g). *In vivo* inhibition of RANKL with the decoy receptor OPG markedly reduced the melanin-producing B16F10 cancer foci and tumour burden in all bones at all time points analysed (Fig. 3d–h). In contrast, the tumour burden and metastasis of B16F10 melanoma cells into ovaries (Supplementary Fig. 7d–f), adrenal glands (Supplementary Fig. 7g–i) and the brain (not shown) were comparable between control and OPG-treated animals. The progressive tumour growth in control vertebrae resulted in spinal cord invasion (Supplementary Fig. 7b) followed by clinical paralysis (Fig. 3i). However, treatment with OPG reduced the tumour burden in vertebrae (Supplementary Fig. 7c), and none of the OPG-treated mice developed clinical paralysis (Fig. 3i). Moreover, control mice injected with

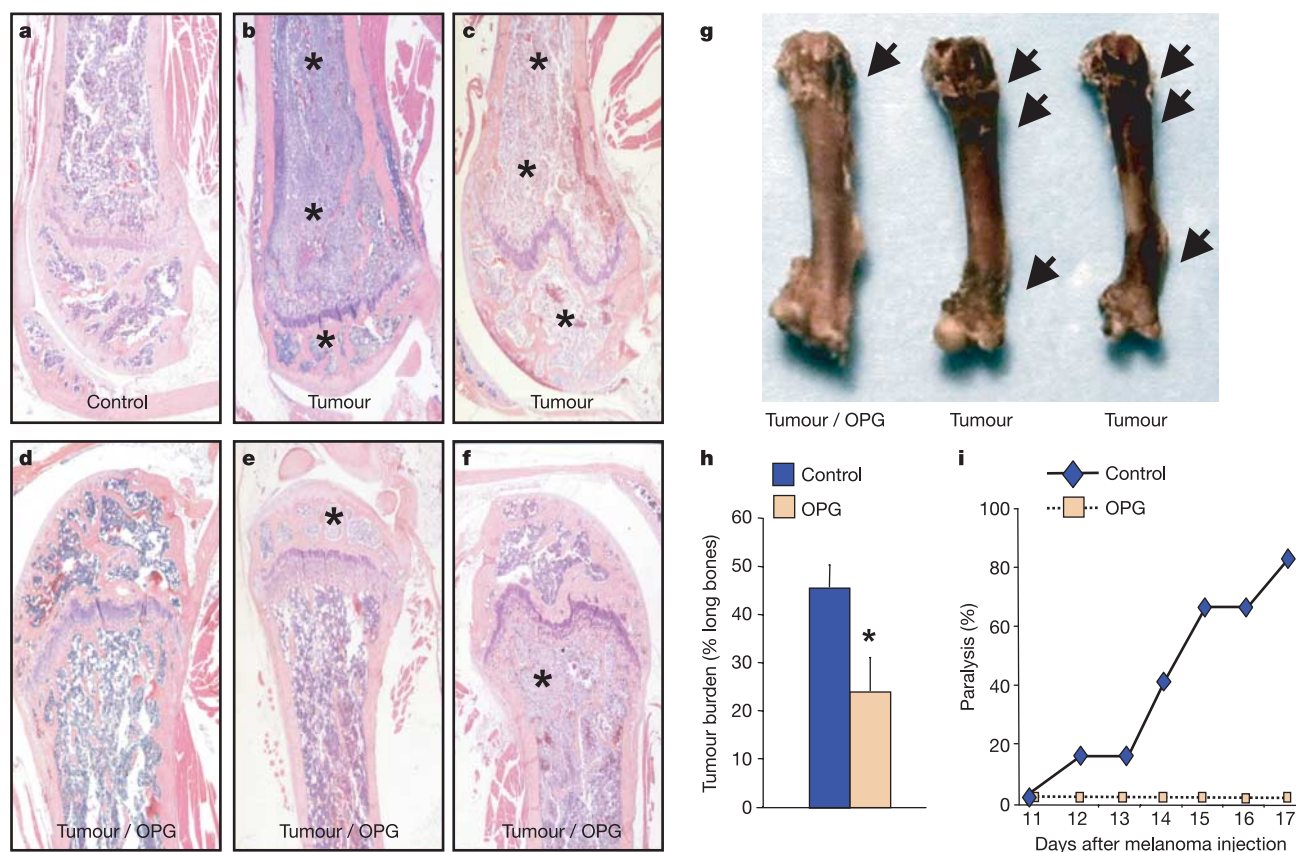


Figure 3 | Inhibition of RANKL/RANK signalling results in reduced tumour metastasis in the bones and abolishes paralysis. a–f, Histology of control long bones (a) and long bone on day 14 (b) or day 17 (c) after injection of B16F10 melanoma cells, and long bones on day 14 (d, e) or day 17 (f) after injection of B16F10 melanoma cells into mice treated with rOPG. Asterisks show typical examples of bone metastases for each treatment. Note that the tumour burden is markedly reduced in OPG-treated mice: no tumour metastasis into long bones (d), tumour foci in the metaphysis but not in other regions of the long bone (e), tumour foci in epiphysis and metaphysis but not in mid-diaphysis (f). Original magnification $\times 5$ for a–f.

g, Macroscopic appearance of long bones on day 14 after injection of B16F10 cells into female recipients. Arrows indicate metastatic foci. Similar results (reduction in the tumour burden of OPG-treated mice) were observed in vertebrae, ribs and skull. **h,** Tumour burden in long bones on day 14 after injection of B16F10 cells ($n = 12$ mice per group). Asterisk, $P < 0.01$ between OPG-treated and untreated groups. Note that all long bones in the control animals showed metastases. The y-axis refers to the average burden ($\pm$ s.d.) of tumour cells in all long bones analysed. **i,** Percentage of mice that developed hind leg paralysis on the indicated days following metastasis of B16F10 cells into vertebrae ($n = 12$ per group).

B16F10 cells showed high morbidity, with some mice dying before the end of the experiment, whereas none of the OPG-treated animals died within the experimental time frame (not shown). Treatment of mice with the bisphosphonate zoledronic acid¹⁸ did not change the tumour burden of B16F10 cancer metastases in bones. Furthermore, using the same experimental system and immunodeficient mice as hosts, the human colon cancer cell lines SW480 and Colo205, which do not express detectable levels of RANK, failed to metastasize into the bones after intracardiac injection ($n = 45$ mice). Thus, *in vivo* inhibition of RANKL with OPG can selectively abrogate metastasis and the tumour burden of B16F10 melanoma cells in bones.

The organ preference of metastatic colonization is influenced by communication between the circulating tumour cells and the target host tissue^{1–4,14}. In particular, osteotropism of certain malignancies is a complication of the primary cancer that often results in severe bone destruction, hypercalcaemia and intractable skeletal pain⁵. In fact, metastases, rather than primary tumours, are responsible for most cancer deaths¹, and it has been estimated that 70% of patients with progressive breast cancer and 84% of prostate cancer patients develop bone metastases^{1,2,19}.

It has long been unclear as to why particular cancers preferentially metastasize to bones. The environment of resorbing bone can provide nutrients to cancer cells, and tumour cells can express osteoclastogenic factors such as parathyroid hormone-related protein (PTHrP) that contribute to local bone degradation and cancer growth^{2,19–21}. Organ-specific chemoattractant molecules have recently been implicated in the preferential homing of breast cancer cell lines to tissues such as lung and lymph nodes^{5,6}. For instance, the chemokine receptor CXCR4 is highly expressed in malignant breast cancer cells, and its ligand, SDF-1 α , is found in organs to which breast cancer frequently metastasizes⁶. However, as inhibition of chemokine receptors *in vivo* only partially blocks the metastatic behaviour of breast cancer cells⁶, other factors must exist to control the tissue-specific migration of epithelial cancer cells.

RANKL is a critical osteoclast differentiation factor that is highly expressed in the bone marrow environment⁸. As we found expression of the receptor RANK on cells from multiple epithelial tumours and a malignant melanoma cell line, which preferentially metastasize to bone, we speculated that RANKL might be one of the long sought-after 'soil' factors⁴ that facilitates metastasis to bone. Our results show that RANKL triggers cytoskeletal changes and migration of several human epithelial tumour cells that express RANK. RANKL also stimulates migration of primary breast epithelial cells and osteoclasts, establishing that RANKL-induced cell migration also occurs in normal, non-transformed cells. Importantly, inhibition of RANKL/RANK signalling by OPG *in vivo* markedly and selectively reduces bone metastasis and tumour burden in a melanoma model that does not activate osteoclasts¹⁵. It remains to be determined whether the dynamics of membrane-bound RANKL and its cleavage to the soluble form present in the plasma of humans and mice could contribute to the metastasis of melanoma cells^{8,11} or metastasis in other model systems²².

In conclusion, our data establish that RANKL can act as a tissue-specific factor for migration of cancer cells and that RANKL is a prominent 'soil' factor for bone-specific metastases of epithelial tumours. Therefore, inhibition of RANKL–RANK interactions may offer a promising therapeutic target for interfering with tumour metastasis and progression in bones.

METHODS

Tumour cell lines. B16F10 murine melanoma cells, MDA-MB-231 human breast cancer, MCF-7 human breast cancer, Hs578T human breast cancer, Colo205 human colon cancer, SW480 human colon cancer, LNCaP human prostate cancer, Du145 human prostate cancer and T47D human epithelial breast tumour cells were used. Non-transformed MCF10A mammary gland epithelial cells and primary mouse mammary gland epithelial cells were freshly isolated from non-pregnant C57BL/6 mouse mammary glands. Animal

experiments were performed in accordance with the guidelines of the Council on Animal Care at the University of Toronto and the University of Western Ontario.

RANKL and RANK expression analysis. Total RNA was isolated from cell lines and mouse tissues using Trizol (Invitrogen), and RANK (*Tnfrsf11a* gene) and RANKL (*Tnfrsf11* gene) mRNA expression were analysed by polymerase chain reaction (PCR). In some experiments, RANK transcripts were confirmed by quantitative real-time PCR with reverse transcription (RT–PCR). RANK mRNA levels were normalized to β -actin levels. Detection of cell-surface expression of RANK protein by fluorescence-activated cell sorting (FACS) used FITC-conjugated human RANKL (amino acids 159–317; Amgen).

RANK signalling, proliferation and cell death assays. Cancer cells were serum-starved for 12 h and then stimulated with recombinant murine RANKL (amino acids 158–316) in the presence or absence of recombinant murine OPG-FC protein (amino acids 22–401; rOPG, both from Amgen)¹¹, SDF-1 α (R&D Systems) or recombinant prolactin (Sigma). In addition, commercially available RANKL (R&D Systems) was used with similar results in osteoclastogenesis, indicating that the observed effects were not attributable to secondary effects of recombinant RANKL (amino acids 158–316).

For western blotting, antibodies reactive to ERK1/ERK2, active ERK1/ERK2 (phosphorylated on Thr 202 and Tyr 204), PKB/AKT, active PKB/AKT (phosphorylated on Ser 473), STAT5A/B, phospho-Stat5A/B (phosphorylated on Tyr 694) (Cell Signalling and Transduction Lab) and actin (Sigma) were used. For actin polymerization studies, tumour cells were stimulated with RANKL or SDF-1 α , and actin polymerization was determined using phalloidin-FITC. Tumour cell proliferation was determined using ³H-thymidine uptake. Cell death was detected by FACS using propidium iodide/AnnexinV-FITC double staining.

Tumour cell migration. Migration of cancer cells was assessed using a 96-well chemotaxis chamber (NeuroProbe Inc.) with fibronectin- (Sigma) coated polycarbonate filters (8- and 12- μ m pore size). All cells were starved for 12 h in DMEM (10 mM HEPES, 0.1% bovine serum albumin), detached using 5 mM EDTA in Ca²⁺/Mg²⁺-free Hank's buffer, counted and resuspended for each assay. rRANKL, rOPG or the chemokines SDF-1 α , 6CKine and CTACK (all chemokines were purchased from R&D Systems) were placed in the lower wells and 5×10^5 B16F10 cells or 2×10^5 human breast, prostate or colon cancer cells were placed in the upper wells. Migration of cells was determined at 37 °C for 16 h (B16F10 cells) or 6 h (human cancer cells) as previously described⁶.

RANK detection on human breast cancer tissue arrays. Paraffin-embedded specimens of tumours, lymph node metastasis, and adjacent normal tissue were collected from 59 female breast cancer patients who underwent surgery between 1988 and 1994, and were analysed retrospectively using protocols approved by the institutional review board of the Medical University of Vienna. Triplicate core biopsies of 0.6 mm were taken from each donor paraffin block and arrayed. Paraffin sections (5- μ m thick) were treated in xylene and rehydrated in a gradient of ethanol. After antigen retrieval by 10 mM sodium citrate (pH 6.0), sections were incubated with a goat polyclonal anti-RANK antibody (M-20, Santa Cruz) for 1 h. Sections were then incubated with biotinylated anti-goat/rabbit IgG antibodies, followed by incubation with streptavidin–peroxidase and 3,3'-diaminobenzidine. Immunostaining was scored on triplicate tissues by two independent observers (T.N. and R.S.) using the following arbitrary scale: 0, no staining; 1, weak staining; 2, medium staining; 3, strong staining. It should be noted that all cancer tissues showed staining in more than 50% of the total tumour area.

In vivo tumour metastasis. Murine B16F10 melanoma cells or human colon cancer cells that do not express RANK were injected into the left cardiac ventricle of 7–10-week-old female C57BL/6 mice or nude mice, respectively⁹. Simultaneously, mice were daily treated with vehicle (PBS), 20 μ g rOPG per day, or zoledronic acid (3 μ g per mouse per day, subcutaneously) as previously described¹⁸. After the final treatment, mice were killed, and bones (femur, tibia, humerus and lumbar vertebrae) and organs (brain, ovary, spleen, kidney and adrenal glands) were collected for histological analysis. Radiographic and histomorphometric analysis of all bones was as previously described²¹. Briefly, tissues were fixed in 10% formalin, sectioned and stained with haematoxylin and eosin to determine the presence of metastases. Midline longitudinal sections of long bones were stained for tartrate-resistant acid phosphatase activity. Two non-serial sections of each bone were assessed. The total tissue section area and the tissue area occupied by tumour cells were measured using the Osteomeasure bone analysis program (Osteometrics Inc.).

Received 11 November; accepted 12 December 2005.

1. Chambers, A. F., Groom, A. C. & MacDonald, I. C. Dissemination and growth of cancer cells in metastatic sites. *Nature Rev. Cancer* 2, 563–572 (2002).

2. Mundy, G. R. Metastasis to bone: causes, consequences and therapeutic opportunities. *Nature Rev. Cancer* **2**, 584–593 (2002).
3. Sloan, E. K. & Anderson, R. L. Genes involved in breast cancer metastasis to bone. *Cell. Mol. Life Sci.* **59**, 1491–1502 (2002).
4. Paget, S. The distribution of secondary growths in cancer of the breast. *Lancet* **1**, 571–572 (1889).
5. Moore, M. A. The role of chemoattraction in cancer metastases. *Bioessays* **23**, 674–676 (2001).
6. Muller, A. *et al.* Involvement of chemokine receptors in breast cancer metastasis. *Nature* **410**, 50–56 (2001).
7. Kong, Y. Y. *et al.* OPGL is a key regulator of osteoclastogenesis, lymphocyte development and lymph-node organogenesis. *Nature* **397**, 315–323 (1999).
8. Lacey, D. L. *et al.* Osteoprotegerin ligand is a cytokine that regulates osteoclast differentiation and activation. *Cell* **93**, 165–176 (1998).
9. Arguello, F., Baggs, R. B. & Frantz, C. N. A murine model of experimental metastasis to bone and bone marrow. *Cancer Res.* **48**, 6876–6881 (1988).
10. Anderson, D. M. *et al.* A homologue of the TNF receptor and its ligand enhance T-cell growth and dendritic-cell function. *Nature* **390**, 175–179 (1997).
11. Teitelbaum, S. L. Bone resorption by osteoclasts. *Science* **289**, 1504–1508 (2000).
12. Fata, J. E. *et al.* The osteoclast differentiation factor osteoprotegerin-ligand is essential for mammary gland development. *Cell* **103**, 41–50 (2000).
13. Verschuere, H. *et al.* Metastatic competence of BW5147 T-lymphoma cell lines is correlated with *in vitro* invasiveness, motility and F-actin content. *J. Leukoc. Biol.* **55**, 552–556 (1994).
14. Liotta, L. A. & Kohn, E. C. The microenvironment of the tumour–host interface. *Nature* **411**, 375–379 (2001).
15. Bakewell, S. J. *et al.* Platelet and osteoclast β_3 integrins are critical for bone metastasis. *Proc. Natl Acad. Sci. USA* **100**, 14205–14210 (2003).
16. Komarova, S. V., Pilkington, M. F., Weidema, A. F., Dixon, S. J. & Sims, S. M. RANK ligand-induced elevation of cytosolic Ca^{2+} accelerates nuclear translocation of nuclear factor κB in osteoclasts. *J. Biol. Chem.* **278**, 8286–8293 (2003).
17. Sanchez-Sweetman, O. H., Lee, J., Orr, F. W. & Singh, G. Direct osteolysis induced by metastatic murine melanoma cells: role of matrix metalloproteinases. *Eur. J. Cancer* **33**, 918–925 (1997).
18. Peyruchaud, O. *et al.* Early detection of bone metastases in a murine model of fluorescent human breast cancer cells: application to the use of the bisphosphonate zoledronic acid in the treatment of osteolytic lesions. *J. Bone Miner. Res.* **16**, 2027–2034 (2001).
19. Guise, T. A. Molecular mechanisms of osteolytic bone metastases. *Cancer* **88** (Suppl.), 2892–2898 (2000).
20. Zhang, J. *et al.* Osteoprotegerin inhibits prostate cancer-induced osteoclastogenesis and prevents prostate tumour growth in the bone. *J. Clin. Invest.* **107**, 1235–1244 (2001).
21. Morony, S. *et al.* Osteoprotegerin inhibits osteolysis and decreases skeletal tumour burden in syngeneic and nude mouse models of experimental bone metastasis. *Cancer Res.* **61**, 4432–4436 (2001).
22. Lelekakis, M. *et al.* A novel orthotopic model of breast cancer metastasis to bone. *Clin. Exp. Metastasis* **17**, 163–170 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements These studies were supported in part by grants from the Canadian Institutes of Health Research (CIHR/IMHA/TAS) and the Canadian Arthritis Network to S.M.S., S.J.D. and S.V.K. We thank W. Boyle, D. Lacey and C. Dunstan for providing rRANKL, RANKL-FITC and rOPG. J.M.P. is supported by the National Cancer Institute of Canada, IMBA, the Austrian National Bank and a European Union Marie Curie Excellence Grant. T.N. holds a European Union Marie Curie Mobility Fellowship.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare competing financial interests: details accompany the paper on www.nature.com/nature. Correspondence and requests for materials should be addressed to J.M.P. (josef.penninger@imba.oeaw.ac.at).

Analysis of a RanGTP-regulated gradient in mitotic somatic cells

Petr Kaláb¹, Arnd Pralle¹, Ehud Y. Isacoff¹, Rebecca Heald¹ & Karsten Weis¹

The RanGTPase cycle provides directionality to nucleocytoplasmic transport, regulating interactions between cargoes and nuclear transport receptors of the importin- β family^{1,2}. The Ran–importin- β system also functions in mitotic spindle assembly and nuclear pore and nuclear envelope formation^{1,3,4}. The common principle underlying these diverse functions throughout the cell cycle is thought to be anisotropy of the distribution of RanGTP (the RanGTP gradient), driven by the chromatin-associated guanine nucleotide exchange factor RCC1 (refs 1, 4, 5). However, the existence and function of a RanGTP gradient during mitosis in cells is unclear. Here we examine the Ran–importin- β system in cells by conventional and fluorescence lifetime microscopy using a biosensor, termed Rango, that increases its fluorescence resonance energy transfer signal when released from importin- β by RanGTP. Rango is predominantly free in mitotic cells, but is further liberated around mitotic chromatin. *In vitro* experiments and modelling show that this localized increase of free cargoes corresponds to changes in RanGTP concentration sufficient to stabilize microtubules in extracts. In cells, the Ran–importin- β –cargo gradient kinetically promotes spindle formation but is largely dispensable once the spindle has been established. Consistent with previous reports^{6–8}, we observe that the Ran system also affects spindle pole formation and chromosome congression *in vivo*. Our results demonstrate that conserved Ran-regulated pathways are involved in multiple, parallel processes required for spindle function, but that their relative contribution differs in chromatin- versus centrosome/kinetochore-driven spindle assembly systems.

To visualize the spatial distribution of the Ran system in living cells, we developed a fluorescence resonance energy transfer (FRET) biosensor termed Rango (Ran-regulated importin- β cargo) that increases its FRET signal when liberated from importin- β by RanGTP (Fig. 1a). Rango contains the importin- β -binding domain (IBB) of human snurportin 1 (ref. 9) flanked by yellow fluorescent protein (EYFP) at the amino terminus and cerulean CFP¹⁰ at the carboxy terminus. In contrast to a sensor based on the IBB of importin- α 1 (ref. 11), Rango displayed little toxicity in somatic cells and did not affect cell cycle progression (data not shown). Upon excitation at 435 nm, Rango exhibited higher emission intensity at the YFP acceptor peak (I_{FRET} at 525 nm) than at the CFP donor wavelength (I_{CFP} at 474 nm), indicative of FRET (Fig. 1b). The ratio of I_{FRET} to I_{CFP} decreased significantly in the presence of importin- β , and this effect was completely reversed by RanGTP, which induced the dissociation of Rango from importin- β (Fig. 1b). In extracts prepared from human HeLa cells, the Rango probe also dynamically reported on the levels of importin- β binding and RanGTP-mediated cargo release (Supplementary Fig. S1).

To quantify the importin- β –cargo interaction using Rango, we measured $I_{\text{FRET}}/I_{\text{CFP}}$ ratio changes in a spectrofluorimeter upon addition of increasing concentrations of importin- β , and plotted the

calculated fractional occupancy of the sensor based on Rango's experimentally determined apparent dissociation constant for importin- β of 2 nM (Fig. 1c; see also Supplementary Fig. S2). At the same time, we measured changes in the fluorescence lifetime of the Rango cerulean donor using fluorescence lifetime imaging microscopy (FLIM; Fig. 1c). As the decrease of the quantum yield due to FRET is accurately reported by a decrease of the donor lifetime (τ_{donor}), FLIM offers a concentration- and cross-bleed-independent FRET detection method that can be used to quantify molecule interactions *in vitro* and in living cells^{12,13}. As expected, the Rango FRET signal decreased with an increase in its fractional occupancy by importin- β , and the average τ_{donor} increased from 2.35 ns to 2.85 ns at 23 °C, and from 2.08 ns to 2.60 ns at 30 °C (Fig. 1c).

Rango introduced into cells by either transient expression or microinjection was efficiently imported into nuclei, where the average τ_{donor} was 2.21 ± 0.07 ns (mean $\pm$ s.d., $N = 10$), indicating that nuclear Rango was almost exclusively free (Fig. 1d). Introduction of the Ran-insensitive importin- β_{71-876} caused the Rango probe to localize to both the cytoplasm and the nucleus, and the average τ_{donor} throughout the cell increased to 2.51 ± 0.05 ns ($N = 7$), similar to the lifetime of a FRET-deficient cerulean control protein (2.59 ± 0.05 ns, $N = 15$), indicating that the increase in Rango sensor lifetime reflected a loss of FRET due to importin- β binding (Fig. 1d; see also Supplementary Fig. S3). We also performed acceptor bleach experiments using confocal laser scanning microscopy that showed that Rango is mostly free in the nucleus (Supplementary Fig. S4), confirming our FLIM analysis. The strong nuclear accumulation of Rango prevented us from analysing its behaviour in the interphase cytoplasm under normal conditions. However, microinjection of low concentrations of wheat germ agglutinin (WGA) partially blocked nuclear transport, causing some Rango to be retained in the cytoplasm, where it displayed lower FRET levels indicative of importin- β binding (Fig. 1e). Thus, Rango enables the RanGTP-dependent disassembly of importin- β –cargo complexes in the nucleus of interphase cells to be visualized.

To measure quantitative differences in Rango binding to importin- β during mitosis, FLIM data were recorded in mitotic HeLa cells transfected with Rango, and Rango's fractional occupancy was estimated based on our *in vitro* calibration data (Fig. 1c). Rango localized throughout the cytoplasm and was largely excluded from mitotic chromosomes (Fig. 2a). In all mitotic HeLa cells in which a gradient could be recorded (36 out of 46 cells; Supplementary Fig. S5), we detected a region of significantly higher FRET (shorter τ_{donor}) surrounding the chromatin ($\tau_{\text{donor}} = 2.21 \pm 0.06$ ns ($N = 36$), corresponding to a Rango–importin- β occupancy of $18 \pm 9\%$) and the FRET signal gradually decreased towards the cell periphery ($\tau_{\text{donor}} = 2.29 \pm 0.07$ ns, $31 \pm 12\%$ occupancy; Fig. 2). Although the FRET values varied considerably between cells (see Supplementary Fig. S5), the average difference of τ_{donor} between

¹Department of Molecular and Cell Biology, University of California, Berkeley, California 94720-3200, USA.

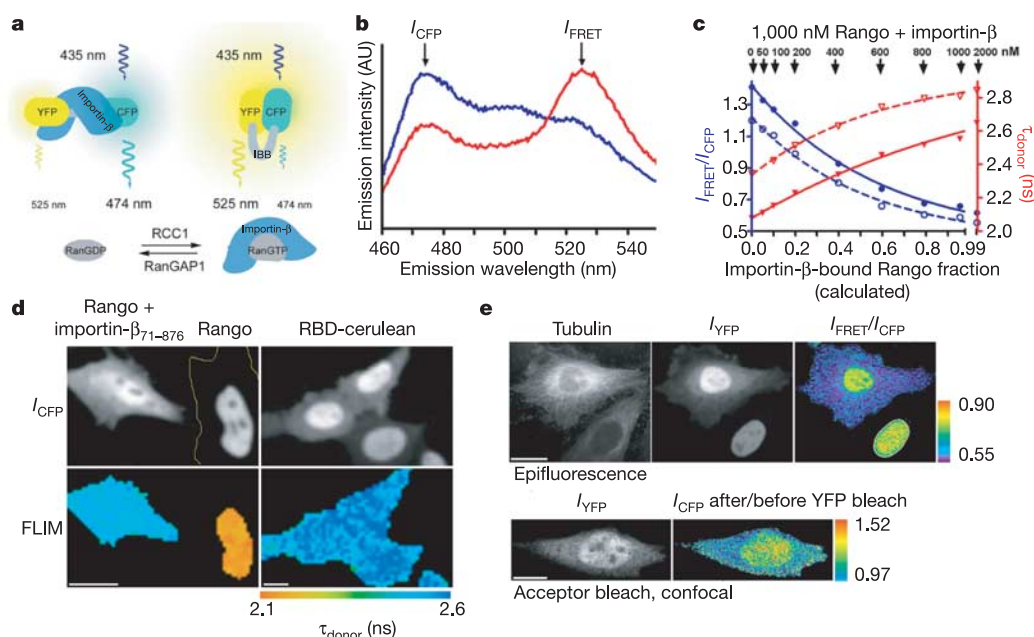


Figure 1 | Characterization of the Rango-importin- β interaction *in vitro*.

a, Schematic of Rango probe. **b**, Emission spectrum of 0.5 μ M Rango excited at 435 nm in the presence of 0.2 μ M RCC1, 1 μ M importin- β , 2 μ M Ran and either 1 mM GDP (blue line) or 1 mM GTP (red line). Arrows indicate I_{CFP} (474 nm) and I_{FRET} (525 nm) emission. AU, arbitrary units. **c**, Rango FRET efficiency determined as $I_{\text{FRET}}/I_{\text{CFP}}$ ratio (blue), or cerulean fluorescence lifetime (red), upon titration of 1,000 nM Rango with 0–2,000 nM importin- β *in vitro* at 23 °C (dashed lines) and at 30 °C (solid lines). The

fractional occupancy of Rango bound to importin- β was calculated.

d, Donor CFP intensity (top) and pseudo-coloured cerulean fluorescence lifetime (bottom) of interphase cells expressing Rango (left panels) or an RBD-cerulean control (right panels). A Rango-expressing cell on the right is outlined; the cell on the left was injected with importin- β_{71-876} .

e, Fluorescence images of cells co-injected with Rango, importin- β , 0.5 mg ml $^{-1}$ WGA and rhodamine-labelled tubulin. Scale bars, 20 μ m.

chromatin and mitotic cytoplasm was 0.08 ± 0.03 ns, corresponding to a $13 \pm 5\%$ decrease in Rango–importin- β binding around chromatin.

These results were also qualitatively confirmed by acceptor photobleaching experiments (Supplementary Figs S6 and S7). The observed gradient did not result from concentration-induced errors in our FLIM measurements, as a modified version of Rango (k-Rango)—which was fused to the DNA-binding domain of the human centromere protein CENP-B 14 —displayed a similar FLIM profile despite a very different localization pattern (Fig. 2a, b). Furthermore, variations in the cargo dissociation constant (in a range between 0.5 and 50 nM) are expected to have only very minor effects on the importin- β occupancy in the mitotic cytoplasm (Supplementary Fig. S8 and data not shown). Thus, our FRET analyses indicate that at equilibrium, high RanGTP concentrations and/or limited importin- β cargo-binding sites exist in the mitotic cytoplasm. Furthermore, they demonstrate the presence of a significant RanGTP-regulated free cargo gradient extending from mitotic chromatin.

To compare embryonic and somatic systems under identical detection conditions, we acquired epifluorescence ratio images in metaphase *Xenopus* egg extracts containing Rango and in HeLa cells microinjected with Rango. Rhodamine-tubulin was used to label microtubules in both cases (Fig. 3a). The dimensions of the cargo gradient were analysed by linescan analysis (Fig. 3b, Methods). Elevated levels of free Rango were observed in *Xenopus* extract spindles in an area extending 15–20 μ m from the chromatin (Fig. 4a), as seen previously with an importin- α 1-based importin- β sensor (YIC) 11 . Although the gradient in the *Xenopus* extract dropped over a greater distance, and was thus significantly less steep than in HeLa cells (3–4 μ m), in both systems it reached to the spindle poles (indicated by asterisks in Fig. 3a, b).

The existence of significant concentrations of free importin- β cargoes throughout the mitotic cytoplasm is inconsistent with simple

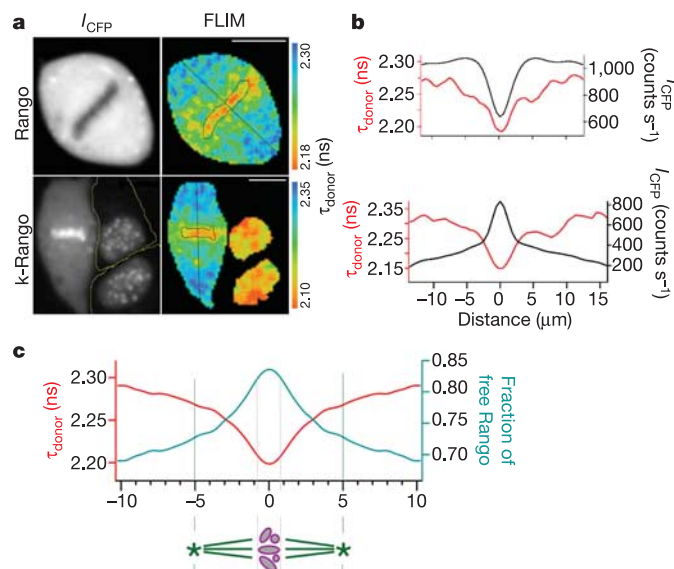


Figure 2 | Detection of the Ran-regulated mitotic Rango gradient in HeLa cells by FLIM.

a, Donor fluorescence (left) and pseudo-coloured FLIM image (right) of a mitotic HeLa cell expressing Rango (top panels) or k-Rango (bottom panels). Chromatin and linescan positions are outlined in the FLIM image. Scale bar, 10 μ m. **b**, Linescan of donor fluorescence lifetime, averaged over 5 μ m (red), and donor intensity (black) obtained from **a**.

c, Average linescans of Rango donor fluorescence lifetime (eight gradients from four HeLa cells, exploiting axial symmetry in the system) (red) and corresponding linescan of unbound Rango fraction (cyan) calculated using the titration curve in Fig. 1c. The average positions of chromatin, mitotic spindle and centrosomes are indicated.

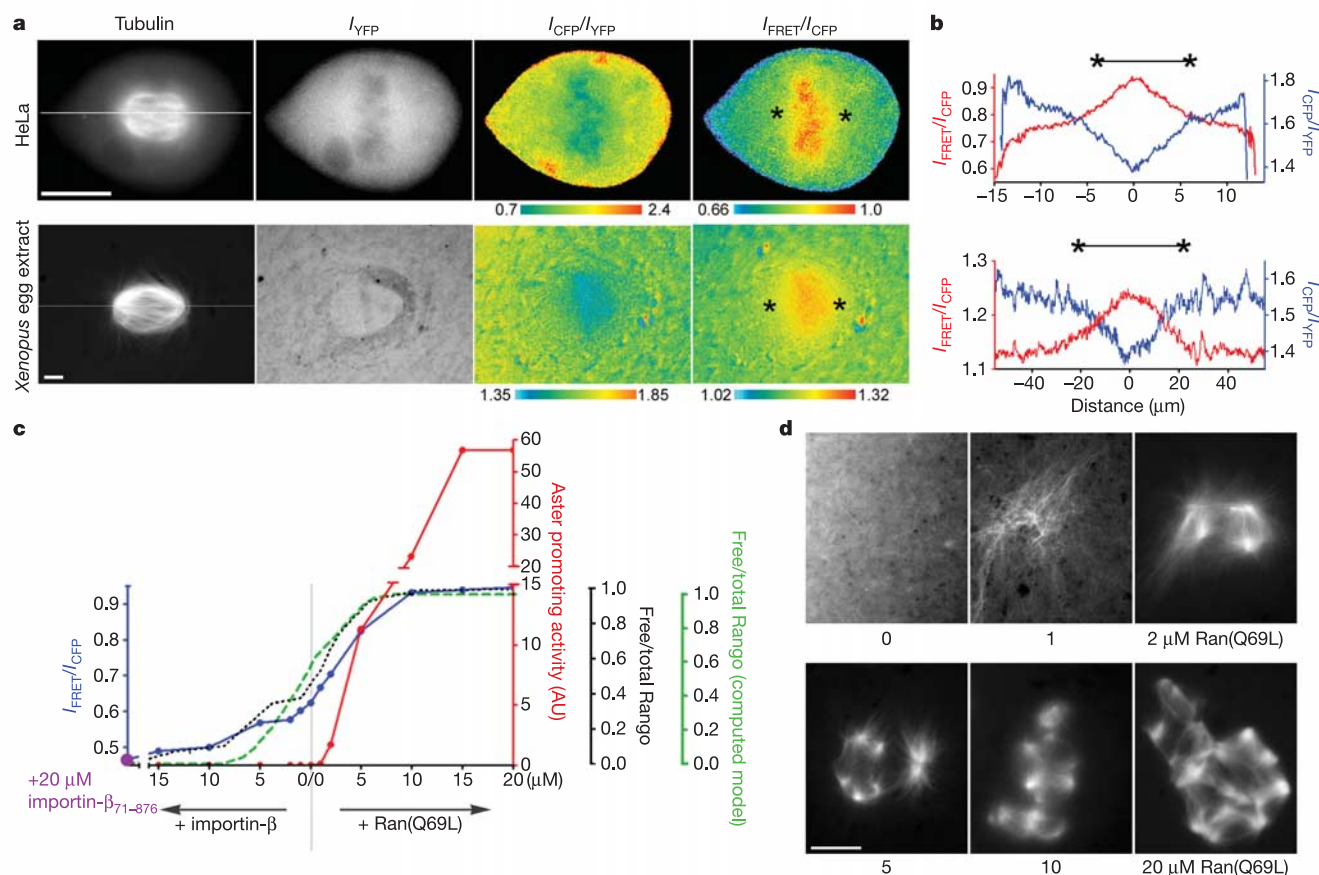


Figure 3 | Comparison of Rango gradient in mitotic HeLa cells and meiotic *X. laevis* egg extracts. **a**, Rango FRET signal in a metaphase HeLa cell microinjected with 18 μM Rango and 1.6 μM rhodamine-tubulin (top), and in *X. laevis* metaphase extract supplemented with 2 μM Rango and 0.5 μM rhodamine-tubulin (bottom). The asterisks indicate the position of spindle poles. The Rango signal in *Xenopus* extract was not background-subtracted, causing an overall shift of the ratio values compared to the HeLa cell. **b**, I_{FRET}/I_{CFP} and I_{CFP}/I_{YFP} ratio linescans (3 μm in HeLa, 5 μm in egg extract) corresponding to the white line in rhodamine-tubulin panels of **a**. **c**, 1 μM Rango and 0.5 μM rhodamine-tubulin was added to mitotic *X. laevis* egg extract, aliquots were supplemented with increasing concentrations of

Ran(Q69L) or importin- β and the I_{FRET}/I_{CFP} ratio was determined by spectrofluorimetry (blue). The fraction of importin- β -free Rango (black) was calculated using *in vitro* titration data (Fig. 1d, see Methods). The response of a minimal computational system (green) was calculated as described in Supplementary Fig. S8. Aster promoting activity (red) was assayed as the average number of mitotic microtubule asters per visual field in fixed samples (AU, see Methods). Note that the scale of aster promoting activity is compressed for values above 20 AU. **d**, Images of rhodamine-tubulin-labelled microtubule structures taken in non-fixed extract samples at the end of analysis. Scale bars, 10 μm .

models that propose complete binding and inhibition of importin- β -regulated activities in the mitotic cytoplasm¹. However, it agrees qualitatively with computer simulations of a minimal Ran system that have been used to calculate free RanGTP concentrations^{15,16} and with our attempts to model importin- β -cargo interactions in cells or extracts (Fig. 3c; see also Supplementary Fig. S8). To investigate whether the observed increase in Ran-regulated cargo liberation around chromatin might regulate microtubule dynamics in mitosis, we performed titration experiments with RanGTP and importin- β in *Xenopus* egg extracts, monitoring, in parallel, the interaction of Rango with importin- β and changes in microtubule morphology (Fig. 3c, d). In the absence of exogenous Ran and importin- β , only $52 \pm 5\%$ ($N = 5$) of Rango was bound to importin- β , and yet the high fraction of free cargoes in the extract did not promote microtubule polymerization (Fig. 3d). However, when the RanGTP concentration was increased by the addition of 1 μM Ran(Q69L), decreasing Rango occupancy by $8 \pm 5\%$ ($N = 5$), bundled microtubules formed. A further increase in RanGTP (2–5 μM Ran(Q69L)) induced formation of relatively large microtubule asters, whereas even higher concentrations (10–30 μM Ran(Q69L)) induced formation of structures with shorter radiating microtubules and more focused centres (Fig. 3d). These results show that the cargo liberation observed around mitotic HeLa chromosomes is quantitatively

similar to the increase in free cargoes sufficient to stimulate microtubule polymerization in extracts. Furthermore, these data suggest that the Ran–importin- β system is poised to respond to small increases in RanGTP concentration from the physiological set point found in the mitotic cytoplasm to regulate microtubule dynamics and organization.

Whereas the Rango FRET signal plateaued at 10 μM of added Ran(Q69L), the number of microtubule asters more than doubled with further increases of Ran(Q69L) to 15–30 μM . This result implies the existence of a class of activity for which regulation in the extracts requires a much higher RanGTP concentration than does a Rango-like cargo. Alternatively, reactions induced by high RanGTP concentrations in the cytoplasmic extracts may mimic conditions of limited diffusion (for example, at the chromatin–cytoplasm interface).

To assess directly the functional significance of the Ran–importin- β system during mitosis in somatic cells, we microinjected a panel of dominant-negative proteins to inhibit the Rango gradient in HeLa cells in either prophase or metaphase. Microinjection of Ran(Q69L) induced the formation of ectopic microtubule asters in the cytoplasm of some cells (Fig. 4b, arrows; see also Supplementary Fig. S9a and Supplementary Table 1) similar to the asters observed in *Xenopus* extracts (Fig. 3d). However, spindles remained intact when metaphase cells were injected with importin- β_{71-876} , a potent dominant-negative

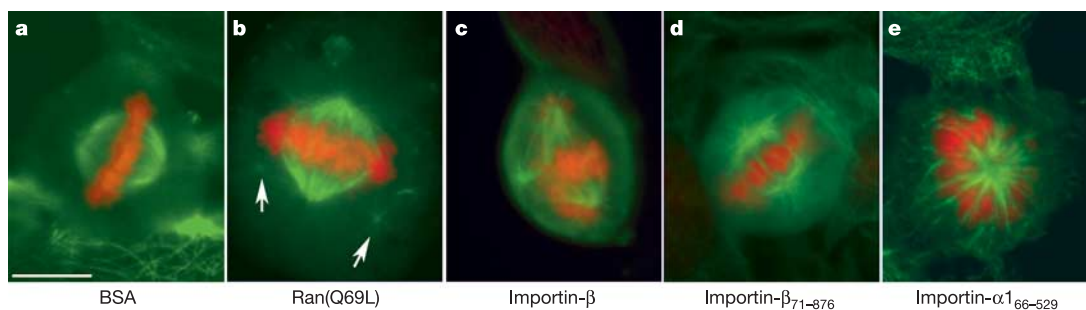


Figure 4 | Mitotic spindle phenotypes induced by Ran system perturbations in somatic cells. **a–e**, Mitotic phenotypes induced by microinjection of BSA (**a**; nonspecific control, 550 μ M in the microinjection needle), Ran(Q69L)

(**b**; 420 μ M), importin- β (**c**; 80 μ M), importin- β_{71-876} (**d**; 420 μ M) and importin- $\alpha_{166-529}$ (**e**; 310 μ M). DNA is coloured red (Hoechst 33342 staining); α -tubulin fluorescence is shown in green. Scale bar, 10 μ m.

inhibitor of chromatin-induced microtubule dynamics and spindle assembly in *Xenopus* extracts^{11,17}. Instead, the most prominent phenotype induced by injection of either importin- β_{71-876} or importin- $\alpha_{166-529}$ in early mitosis was a delay at the prometaphase to metaphase transition, frequently associated with monoastal microtubule arrays (Fig. 4d, e; see also Supplementary Fig. S9 and Supplementary Table 1). The induction of large monoastal microtubule structures indicates that the normal progression from a radial to bipolar microtubule arrangement during prophase requires an intact cargo gradient in cells. This conclusion is consistent with modelling studies¹⁸ and suggests that Ran-gradient-regulated stabilization of microtubules around chromatin supports a search and capture mechanism of microtubule-kinetochore attachment. Notably, in addition to a delay in prometaphase, injection of full-length importin- β also induced split spindle poles (Fig. 4c; see also Supplementary Table 1), consistent with a model that Ran and importin- β may function in the regulation of centrosomes⁶.

Our results suggest that the RanGTP gradient provides a significant kinetic advantage during the early stages of spindle assembly in primarily centrosome-driven somatic cells. However, in contrast to the situation for extracts, once a bipolar spindle is established in cells, the RanGTP gradient and the Ran-importin- β cargo regulation appears to be largely dispensable for spindle integrity. This indicates that in cells, mitotic spindles are built and maintained by multiple, parallel pathways, and demonstrates that centrosome/kinetochore- and chromatin-driven systems differentially use Ran and importin- β to promote mitotic spindle assembly. Notably, in both HeLa cells and *Xenopus* egg extracts, the steepness of the Ran-regulated gradient seems to be adjusted to the enormous differences in the spindle size (Fig. 3a, b). This organization permits relative differences in cargo occupancy to influence events between the spindle poles and chromatin in both systems.

On the basis of our results, we propose that the mitotic cytoplasm operates near a physiological threshold in which positive and negative regulators are at equilibrium. Such a system would be poised to break the threshold in response to small local changes in RanGTP concentration and, for example, influence microtubule stability around chromatin in prophase cells. This behaviour may allow the Ran-importin- β pathway to locally regulate its targets and to signal both chromatin- and centrosome-driven events in mitosis.

METHODS

Cloning and protein expression. A description of all the clones used in this study and details of protein expression are given in the Supplementary Methods. **Fractional occupancy of Rango titrated with importin- β .** Rango-importin- β fractional occupancy (Fig. 1c) was calculated as follows:

$$K_d = \beta C / C_{\beta C}$$

$$K_d = (\beta_T - C_{\beta C})(C_T - C_{\beta C}) / C_{\beta C}$$

where $C_{\beta C}$ is the concentration of the importin- β -cargo complex (fractional occupancy), β is the concentration of free importin- β , β_T is the total concentration of importin- β , C is the concentration of free importin- β cargo, C_T is the total concentration of importin- β cargo and K_d is the dissociation constant, resulting in $C_{\beta C} = 0.5(C_T + \beta_T + K_d - \sqrt{(C_T + \beta_T + K_d)^2 - 4\beta_T C_T})$.

Spectrophotometry. Emission spectra were analysed with a Fluorolog 2 spectrofluorimeter controlled by Datamax 2.2 (Jobin Yvon Spex) and the Grams 3.04 II software package (Galactic Industries). Details are given in the Supplementary Methods.

Cell culture and transfection. BHK21 cells and HeLa cells were purchased from ATCC. tsBN2 cells were a gift of T. Nishimoto and M. Dasso. Cells were maintained in Opti-MEM (Gibco, Invitrogen) with 4% fetal bovine serum at 37 °C, 5% CO₂, except for tsBN2 cells, which were kept at 33 °C, 5% CO₂. For cell transfection, Eugene 6 (Roche Diagnostics) was used according to the manufacturer's protocol.

Microinjection and immunofluorescence. Cells were microinjected using an Olympus IX71 microscope equipped with a FemtoJet microinjector (Eppendorf), and analysed by immunofluorescence to visualize microtubules and DNA using an Olympus BX51 microscope equipped with a Hamamatsu CA 742-98 CCD camera. Details are given in the Supplementary Methods.

Live cell epifluorescence imaging. Live cell epifluorescence ratio imaging was performed with a Nikon E600 microscope equipped with a Hamamatsu C4742-98 CCD camera as described previously¹¹. Additional details are given in the Supplementary Methods.

Fluorescence lifetime and confocal microscopy. Data sets of spatially resolved, time-correlated single photon counting (TCSPC) were acquired on an inverted Zeiss LSM510 Axiovert 200M microscope equipped with a TCSPC controller (Becker & Hickl SPC-730). Confocal microscopy was performed with a Zeiss LSM 510 META laser scanning confocal microscope. Additional details are provided in the Supplementary Methods.

***Xenopus laevis* egg extracts.** Assays for the detection of the Rango I_{FRET}/I_{CFP} signal during mitotic spindle assembly in *X. laevis* egg extracts were performed as described previously¹¹ with rhodamine-tubulin and 2 μ M Rango in the extract instead of YIC. Details are given in the Supplementary Methods.

Statistical analyses. Statistical analyses were performed with Excel (Microsoft) and with GraphPad Prism version 4.00 for Windows, GraphPad Software (<http://www.graphpad.com>).

Received 18 November 2005; accepted 18 January 2006.

1. Weis, K. Regulating access to the genome: nucleocytoplasmic transport throughout the cell cycle. *Cell* **112**, 441–451 (2003).
2. Pemberton, L. F. & Paschal, B. M. Mechanisms of receptor-mediated nuclear import and nuclear export. *Traffic* **6**, 187–198 (2005).
3. Hetzer, M., Gruss, O. J. & Mattaj, I. W. The Ran GTPase as a marker of chromosome position in spindle formation and nuclear envelope assembly. *Nature Cell Biol.* **4**, E177–E184 (2002).
4. Harel, A. & Forbes, D. J. Importin- β : conducting a much larger cellular symphony. *Mol. Cell* **16**, 319–330 (2004).
5. Hetzer, M., Bilbao-Cortes, D., Walther, T. C., Gruss, O. J. & Mattaj, I. W. GTP hydrolysis by Ran is required for nuclear envelope assembly. *Mol. Cell* **5**, 1013–1024 (2000).
6. Ciciarello, M. et al. Importin- β is transported to spindle poles during mitosis and regulates Ran-dependent spindle assembly factors in mammalian cells. *J. Cell Sci.* **117**, 6511–6522 (2004).
7. Arnautov, A. & Dasso, M. The Ran GTPase regulates kinetochore function. *Dev. Cell* **5**, 99–111 (2003).

8. Arnaoutov, A. *et al.* Crm1 is a mitotic effector of Ran-GTP in somatic cells. *Nature Cell Biol.* **7**, 626–632 (2005).
9. Huber, J., Dickmanns, A. & Luhrmann, R. The importin- β binding domain of snurportin1 is responsible for the Ran- and energy-independent nuclear import of spliceosomal U snRNPs *in vitro*. *J. Cell Biol.* **156**, 467–479 (2002).
10. Rizzo, M. A., Springer, G. H., Granada, B. & Piston, D. W. An improved cyan fluorescent protein variant useful for FRET. *Nature Biotechnol.* **22**, 445–449 (2004).
11. Kalab, P., Weis, K. & Heald, R. Visualization of a Ran-GTP gradient in interphase and mitotic *Xenopus* egg extracts. *Science* **295**, 2452–2456 (2002).
12. Suhling, K., French, P. M. & Phillips, D. Time-resolved fluorescence microscopy. *Photochem. Photobiol. Sci.* **4**, 13–22 (2005).
13. Becker, W. *et al.* Fluorescence lifetime imaging by time-correlated single-photon counting. *Microsc. Res. Tech.* **63**, 58–66 (2004).
14. Shelby, R. D., Hahn, K. M. & Sullivan, K. F. Dynamic elastic behaviour of alpha-satellite DNA domains visualized *in situ* in living human cells. *J. Cell Biol.* **135**, 545–557 (1996).
15. Gorlich, D., Seewald, M. J. & Ribbeck, K. Characterization of Ran-driven cargo transport and the RanGTPase system by kinetic measurements and computer simulation. *EMBO J.* **22**, 1088–1100 (2003).
16. Riddick, G. & Macara, I. G. A systems analysis of importin- α - β mediated nuclear protein import. *J. Cell Biol.* **168**, 1027–1038 (2005).
17. Nachury, M. V. *et al.* Importin- β is a mitotic target of the small GTPase Ran in spindle assembly. *Cell* **104**, 95–106 (2001).
18. Wollman, R. *et al.* Efficient chromosome capture requires a bias in the 'search-and-capture' process during mitotic-spindle assembly. *Curr. Biol.* **15**, 828–832 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors wish to thank T. Nishimoto, M. Dasso, J. Fang, M. A. Rizzo, D. W. Piston and F. Melchior for providing reagents, and C. Weirich for performing fluorescence polarization assays. We are grateful to A. Arnaoutov for discussion and sharing unpublished results, C. Weirich, M. Blower, A. Madrid and H. Aaron for critical reading of the manuscript, and members of the Heald and Weis laboratories for discussions. The research described in this article was supported in part by Philip Morris USA Inc. and Philip Morris International (R.H.), and by grants from the National Institute of Health (E.Y.I., R.H. and K.W.).

Author Contributions P.K. and A.P. contributed equally to this project.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.H. (heald@socrates.berkeley.edu) or K.W. (kweis@berkeley.edu).

LETTERS

Oncogenic activity of Cdc6 through repression of the *INK4/ARF* locus

Susana Gonzalez¹, Peter Klatt¹, Sonia Delgado⁴, Esther Conde², Fernando Lopez-Rios⁵,
Montserrat Sanchez-Céspedes², Juan Mendez³, Francisco Antequera⁴ & Manuel Serrano¹

The *INK4/ARF* locus encodes three tumour suppressors (p15^{INK4b}, ARF and p16^{INK4a}) and is among the most frequently inactivated loci in human cancer^{1,2}. However, little is known about the mechanisms that govern the expression of this locus. Here we have identified a putative DNA replication origin at the *INK4/ARF* locus that assembles a multiprotein complex containing Cdc6, Orc2 and MCMs, and that coincides with a conserved noncoding DNA element (regulatory domain RD^{INK4/ARF}). Targeted and localized RNA-interference-induced heterochromatinization of RD^{INK4/ARF} results in transcriptional repression of the locus, revealing that RD^{INK4/ARF} is a relevant transcriptional regulatory element. Cdc6 is overexpressed in human cancers, where it might have roles in addition to DNA replication^{3–5}. We have found that high levels of Cdc6 result in RD^{INK4/ARF}-dependent transcriptional repression, recruitment of histone deacetylases and heterochromatinization of the *INK4/ARF* locus, and a concomitant decrease in the expression of the three tumour suppressors encoded by this locus. This mechanism is reminiscent of the silencing of the mating-type *HM* loci in yeast by replication factors⁶. Consistent with its ability to repress the *INK4/ARF* locus, Cdc6 has cellular immortalization activity and neoplastic transformation capacity in cooperation with oncogenic Ras. Furthermore, human lung carcinomas with high levels of Cdc6 are associated with low levels of p16^{INK4a}. We conclude that aberrant expression of Cdc6 is oncogenic by directly repressing the *INK4/ARF* locus through the RD^{INK4/ARF} element.

The identification of regulatory elements is challenging; in some instances, regulatory elements have been found at, or in proximity to, replication origins^{7–9}. We have searched for replication initiation sites at the *INK4/ARF* locus by measuring nascent-strand abundance along the locus in two human cell lines: embryo kidney HEK-293T and astrocytoma GO-G-UVW cells. A putative replication origin was found 1.5 kilobases (kb) upstream of the ATG start codon of p15^{INK4b} in the two cell lines (Fig. 1a and Supplementary Fig. 1). The location of the replication origin coincides with a DNA element conserved among mammalian *INK4/ARF* loci (Supplementary Fig. 2). Specifically, this conserved element spans over ~350 base pairs (bp) with more than 60% identity, including a shorter segment of ~150 bp with more than 80% identity between mammals (Supplementary Fig. 2c). The sequence requirements of mammalian replication origins are relaxed and do not possess identifiable conserved sequence elements⁹, whereas transcriptional regulatory elements are often conserved⁷. On the basis of the conservation of the *INK4/ARF* putative replication origin, we hypothesized that it could also display transcriptional regulatory activity. In a first approximation, a fragment containing this region was found to enhance ($\geq$ fourfold) the activity of a reporter gene under the control

of an SV40 minimal promoter in an orientation-independent and copy-number-dependent manner (Supplementary Fig. 3). The above observations suggest that the putative replication origin at the *INK4/ARF* locus may possess transcriptional regulatory activity and, therefore, we have named it regulatory domain (RD^{INK4/ARF}).

RNA interference (RNAi) machinery, in addition to degrading complementary messenger RNAs, can induce the heterochromatinization of complementary genomic DNA regions^{10,11}. We have used this tool to test the relevance of RD^{INK4/ARF} in its natural genomic context. A pool of short interfering (si)RNAs, or their derived retroviral constructs expressing short-hairpin (sh)RNAs, were targeted to human RD^{INK4/ARF} (hRD) in kidney HEK-293T cells and IMR90 fibroblasts, and to murine RD^{INK4/ARF} (mRD) in mouse embryo fibroblasts (MEFs). Heterochromatinization was examined by measuring the presence of trimethylated lysine 9 on histone H3 (H3K9me3) at RD^{INK4/ARF} by chromatin immunoprecipitation (ChIP). The amount of H3K9me3 at RD^{INK4/ARF} increased as a result of the presence of siRNA-RD or shRNA-RD, thus indicating RNAi-induced heterochromatinization (Fig. 1b). This effect was not observed when we examined the intron of *INK4b* (Fig. 1b) or a non-related genomic region, such as the *p73* gene (not shown). Notably, the presence of RNAi targeted to RD^{INK4/ARF} strongly reduced the levels of the three mRNAs and corresponding proteins encoded by the locus, namely, p15^{INK4b}, ARF and p16^{INK4a} (Fig. 1b). Mutant shRNAs that were not perfectly complementary to RD^{INK4/ARF} had no effect on RD^{INK4/ARF} heterochromatinization nor on p16^{INK4a} levels (Supplementary Fig. 4). Furthermore, when siRNAs were directed against a different genomic element of the locus, such as the *INK4a* promoter, we only observed repression of p16^{INK4a}, but not of ARF (data not shown). To confirm and extend the above data, introduction of shRNA-mRD into primary wild-type MEFs recapitulated the immortalization and neoplastic transformation phenotypes of *INK4a/ARF*-null MEFs¹² (lacking exons 2 and 3; for a map see Supplementary Fig. 2a), as evaluated by colony formation assays (Fig. 1c) and oncogenic cooperation assays with Ras (Fig. 1d). Together, these results demonstrate that the functionality of RD^{INK4/ARF} is critical for the transcriptional activity of the *INK4/ARF* locus.

There are numerous examples of coordinated interaction between replication and transcriptional regulation¹³ and, on this basis, we hypothesized that replication factors might have a dual role at RD^{INK4/ARF}. We focused on Cdc6 because it is aberrantly overexpressed in some human cancers^{3–5}. Consistent with the role of RD^{INK4/ARF} as a putative replication origin, specific binding of epitope-tagged Cdc6 to RD^{INK4/ARF}, but not to neighbouring regions, was observed in a variety of human cells (see Fig. 2a for HEK-293T cells; similar data for IMR90 and osteosarcoma SAOS2 cells are not

¹Tumor Suppression Group, ²Lung Cancer Group, and ³DNA Replication Group, Spanish National Cancer Research Center (CNIO), E-28029 Madrid, Spain. ⁴Instituto de Microbiología Bioquímica, CSIC/Universidad de Salamanca, E-37007 Salamanca, Spain. ⁵Departamento de Anatomía Patológica, Hospital Universitario "12 de Octubre", E-28041 Madrid, Spain.

shown). In these experiments, ectopic expression of Cdc6 resulted in moderate overexpression (~fivefold relative to normal levels; Fig. 2a, see also Supplementary Fig. 8c), within the range observed in human tumours³. As a positive control, we also detected binding of ectopic Cdc6 to the well-characterized human lamin B2 replication origin¹⁴ (data not shown). Notably, endogenous Cdc6 was also observed associated to RD^{INK4/ARF}, but not to the *INK4b* intron or *p73*, and this interaction was disrupted by the presence of siRNA-hRD (Fig. 2b). As further confirmation of the assembly of a replication complex at RD^{INK4/ARF}, we found site-specific binding of Orc2 (Fig. 2a) and Cdc6-dependent loading and spreading of endogenous MCMs throughout the *INK4/ARF* locus (Supplementary Fig. 5), in agreement with current views on Cdc6 function¹⁵. As an additional control, mutant Cdc6(D284A/E285A) in the Walker-B motif conserved in DNA-dependent ATPases was partially defective in binding RD^{INK4/ARF} and was unable to load MCMs (Supplementary

Figs 5 and 6), as predicted from the involvement of this domain in Cdc6-mediated MCM loading¹⁶. These results indicate that the regulatory element RD^{INK4/ARF} assembles a multiprotein complex that includes the cancer-associated replication factor Cdc6.

Next, we studied the effect of high Cdc6 levels on the expression of the *INK4/ARF* locus. As shown in Fig. 2c, increased Cdc6 in HEK-293T cells leads to a substantial reduction in the expression of the three *INK4/ARF*-encoded genes (similar data were also obtained in MEFs, Supplementary Fig. 10a, and in IMR90 cells, data not shown). A role of Cdc6 in transcriptional repression through the RD^{INK4/ARF} element was further supported with reporter assays using constructs harbouring human or murine RD^{INK4/ARF} (Fig. 2d). In these assays, the above-mentioned Walker-B Cdc6 mutant was completely inactive as a repressor (Supplementary Fig. 6; see in the same figure the analysis of additional Cdc6 mutants). Finally, we wondered whether the repressive effect of Cdc6 on the *INK4/ARF*

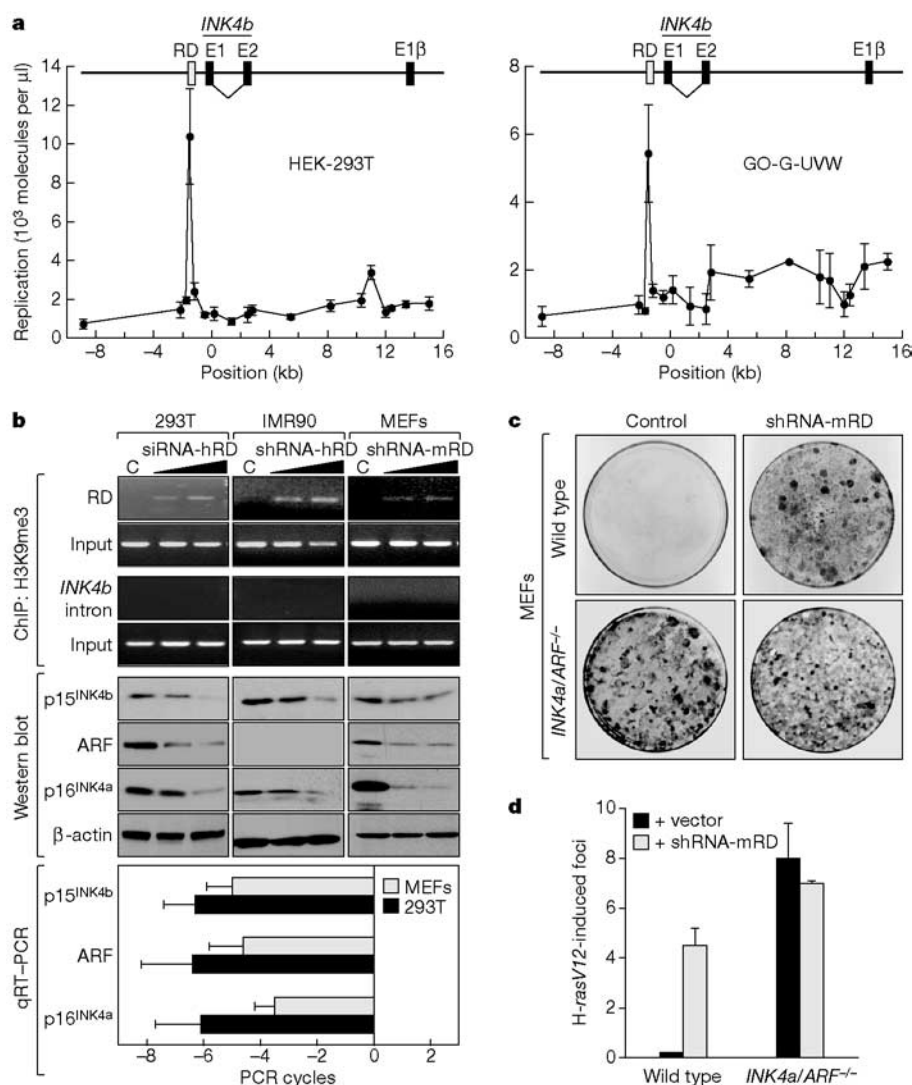


Figure 1 | Identification of a regulatory domain, RD^{INK4/ARF}, in the *INK4/ARF* locus. **a**, Localization of a putative replication origin in the *INK4/ARF* locus by competitive PCR of nascent DNA strands (for details see Supplementary Fig. 1 and Supplementary Table 1). Position 0 corresponds to the ATG of *p15^{INK4b}*. **b**, RNAi was produced by transient transfection of siRNA-hRD in HEK-293T cells, or by stable retroviral transduction of shRNA-hRD in IMR90 cells or shRNA-mRD in primary wild-type MEFs. Heterochromatinization was evaluated by ChIP against H3K9me3. As controls (C), we used siRNA-luciferase (in the case of HEK-293T cells) or empty vector (in the case of IMR90 and MEFs). In agreement with previous results²⁸, ARF could not be detected in IMR90 cells. Transcripts were

quantified in cells expressing the highest amount of the corresponding RNAi used in the upper part of the panel (see Methods). Assays were performed after 48 h in the case of siRNA transfection, or 72 h post-selection of shRNA-transduced cells. **c**, Colony formation assay using primary MEFs retrovirally transduced with shRNA-mRD or an empty vector (control). **d**, Foci formation in primary MEFs (10⁶ cells) retrovirally transduced with shRNA-mRD or an empty vector (control) and subsequently transfected with a plasmid encoding oncogenic Ras (10 μg). The figure shows the average and s.d. of two independent assays. All the data shown in **b**, **c** and **d** are representative of at least two independent assays.

locus was general to other loci containing well-characterized replication origins, such as the loci encoding c-Myc¹⁷, Dnmt1 (ref. 18) and Mcm4 (ref. 19). In contrast to the observed repressive effect on the *INK4/ARF* locus, overexpression of Cdc6 had no effect on the levels of the above-mentioned proteins (Supplementary Fig. 7), suggesting that the repressive effect of Cdc6 is not widespread and does not affect every gene located in the proximity of a replication origin.

Histone deacetylation has been identified, both in yeast and vertebrates, as the earliest histone alteration associated with gene silencing^{20,21}. We reasoned that Cdc6 overexpression could recruit histone deacetylases at the *INK4/ARF* locus. ChIP assays using antibodies specific for either histone deacetylase 1 or 2 (HDAC1 or HDAC2) indicated that overexpression of Cdc6 caused an increase in

the amount of these proteins at RD^{*INK4/ARF*}, as well as at the *ARF* and *INK4a* promoters (Fig. 2e). The recruitment of HDACs at these sites correlated with a decrease in the acetylation of histones H3 and H4 (Supplementary Fig. 8a). As additional controls, the presence of the histone deacetylase inhibitor trichostatin A prevented Cdc6-induced deacetylation of the *INK4/ARF* locus (Supplementary Fig. 8b), and basal levels of acetylated H3 did not change during the cell cycle (Supplementary Fig. 9). Finally, we examined the stability of Cdc6-induced chromatin changes, as well as the appearance of heterochromatin marks. After two weeks of Cdc6 overexpression in HEK-293T cells, HDACs were still present at the *INK4/ARF* locus, and there was an increase in H3K9me3, suggesting Cdc6-triggered heterochromatinization (Supplementary Fig. 8c; similar data were also obtained with MEFs, see Supplementary Fig. 11). We conclude that high levels of Cdc6 are capable of specifically repressing the *INK4/ARF* locus through a mechanism that implies the recruitment of histone deacetylases and the induction of heterochromatinization.

Following on from the above observations, we investigated whether Cdc6 could recapitulate the immortalization and neoplastic transformation phenotypes of *INK4a/ARF*^{-/-} MEFs¹². Colony formation analyses showed a significant increase in colonies in primary wild-type MEFs induced by ectopic expression of Cdc6 (Fig. 3a). In addition, Cdc6 cooperated with oncogenic Ras when introduced into primary wild-type MEFs, as assessed by the generation of neoplastic foci (Fig. 3b; which were able to form tumours in nude mice, data not shown) and by the ability to proliferate in soft agar (Fig. 3c). The immortalization and oncogenic activities of Cdc6 were not noticeable in *INK4a/ARF*^{-/-} MEFs, suggesting that this locus is a critical mediator of the oncogenic activity of Cdc6. Moreover,

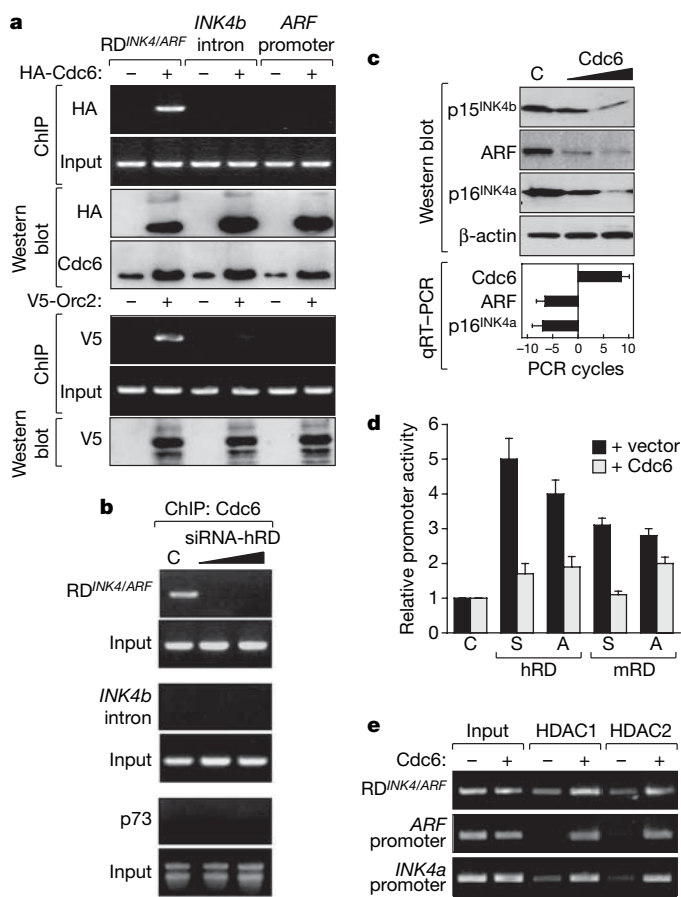


Figure 2 | Binding of Cdc6 to RD^{*INK4/ARF*} and repression of the *INK4/ARF* locus. **a**, Site-specific loading of Cdc6 and Orc2 to RD^{*INK4/ARF*}. Assays were performed in HEK-293T cells 48 h after transient transfection of the indicated proteins. **b**, Binding of endogenous Cdc6 to RD^{*INK4/ARF*} and abrogation of Cdc6 binding by RNAi-induced heterochromatinization in HEK-293T cells (for details see Fig. 1b). **c**, High levels of Cdc6 repress the expression of the *INK4/ARF* locus. HEK-293T cells were transiently transfected with increasing amounts of Cdc6 and analysed 72 h later by western blot or RT-PCR (only for the highest amount of Cdc6 transfected). The higher amount of transfected Cdc6 corresponds to the amount transfected in **a** and in Supplementary Fig. 8c (see Methods). **d**, Inhibition of RD^{*INK4/ARF*} enhancer activity by Cdc6. Relative luciferase activity in HeLa cells co-transfected with or without Cdc6, along with a luciferase reporter driven by a minimal SV40 promoter alone (C) or containing the human or murine RD^{*INK4/ARF*} (hRD or mRD, respectively) in sense (S) or antisense (A) orientation. Assays were performed 48 h after transfection. Values represent mean $\pm$ s.d. ($n = 3$). **e**, ChIP assays were performed 72 h after transient transfection of HEK-293T cells using antibodies against HDAC1 or HDAC2. All the data shown are representative of a minimum of two independent assays.

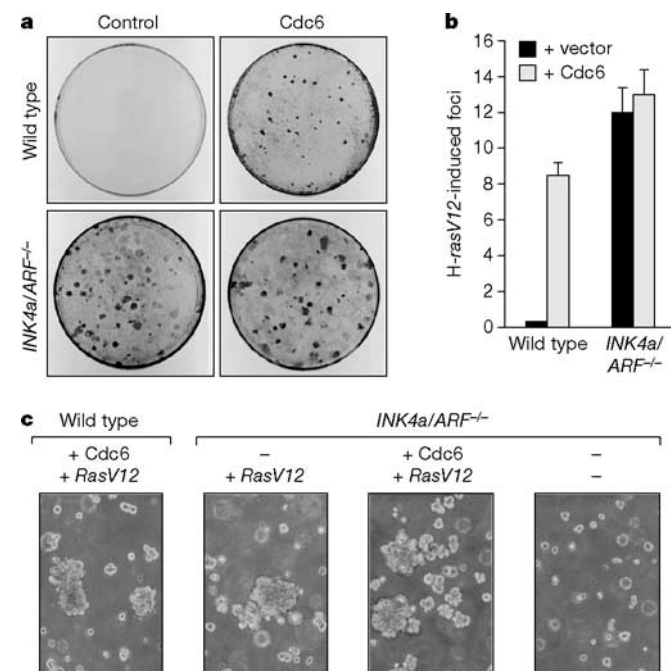


Figure 3 | Oncogenic activity of Cdc6. **a**, Colony formation assay using primary wild-type and *INK4a/ARF*^{-/-} MEFs infected with Cdc6, or an empty vector (control). **b**, Foci formation in wild-type and *INK4a/ARF*^{-/-} MEFs (10⁶ cells) transfected with a plasmid encoding oncogenic Ras (10 μ g) together with the same amount of a plasmid expressing Cdc6, or an empty vector (control). The figure shows the average and s.d. of two independent assays. **c**, Proliferation in soft agar of primary MEFs expressing Cdc6 and/or oncogenic Ras. Cultures of primary cells were retrovirally transduced with Cdc6 (or empty vector) and then transfected with oncogenic Ras (10 μ g). As a control, primary *INK4a/ARF*^{-/-} MEFs were not able to proliferate in soft agar.

overexpression of Cdc6 had no detectable effects on the cell cycle or proliferation rate of primary wild-type MEFs (Supplementary Fig. 10). Together, these observations support the concept that the main effect of Cdc6 overexpression is not on proliferation *per se*, but rather on the suppression of the *INK4/ARF*-dependent barriers to immortalization and oncogenic transformation. Interestingly, Ras-transformed *INK4a/ARF*^{-/-} MEFs had a normal, basal amount of H3K9me3 at the *INK4/ARF* locus, whereas Ras/Cdc6-transformed wild-type MEFs had increased levels of H3K9me3 (Supplementary Fig. 11), thus extending the association between Cdc6 overexpression and *INK4/ARF* heterochromatinization to the context of neoplastic transformation.

To determine the relevance of the above findings in human tumours, we studied the relationship between the protein levels of Cdc6 and p16^{INK4a} in non-small-cell lung carcinomas (NSCLCs; *n* = 162). Following previously described criteria³, tumours were classified as Cdc6-low (normal levels) or Cdc6-high (abnormally high levels). The levels of Cdc6 did not correlate with the proliferation index of the tumours (Fig. 4a, see data for proliferation marker Ki67), which is in agreement with previous reports³ and with our current observations (see above and Supplementary Fig. 10). On the other hand, tumours were also categorized as p16-negative (complete absence of nuclear immunostaining), p16-low (1–25% of positive nuclei) or p16-high (>25% of positive nuclei). Tumours classified as p16-negative were excluded from subsequent analysis because the underlying cause for the absence of p16 could be due to

genetic deletion or promoter methylation of the locus, which are frequent alterations in NSCLCs (50–70%)²². Of note, among those tumours retaining expression of p16, there was a reciprocal association between Cdc6 and p16^{INK4a} expression levels in NSCLCs (Fig. 4a). These observations further support the concept that overexpression of Cdc6 is oncogenic through downregulation of the *INK4/ARF* locus.

Our data are compatible with a mechanistic model by which the *INK4/ARF* locus is positively governed by a conserved DNA regulatory domain (RD^{*INK4/ARF*}) (Fig. 4b). This regulatory domain is sensitive to the levels of Cdc6 in such a manner that increased levels of Cdc6 result in recruitment of heterochromatinizing activities and downregulation of the three tumour suppressors encoded by the *INK4/ARF* locus (Fig. 4b). This model, although unprecedented in vertebrates, is remarkably similar to the silencing of the mating-type *HM* loci of the yeast *Saccharomyces cerevisiae* through a multi-protein complex that contains replication factors⁶. The oncogenic mechanism reported here for Cdc6 may constitute a relevant alternative pathway for the functional inactivation of the *INK4/ARF* locus in human cancer.

METHODS

Nascent-strand isolation and PCR-based origin localization assay. Exponentially growing HEK-293T or GO-G-UVW cells were lysed and overlaid directly on top of a seven-step alkaline sucrose gradient and centrifuged as previously described²³. DNA from fractions containing nascent strands between 1 kb and 3 kb was used for quantitative PCR. Eighteen pairs of primers and the corresponding sets of competitors (Supplementary Table 1) across a 25-kb region spanning the *INK4b/ARF* genes were used to measure the amount of nascent strands by competitive PCR²³.

Cells and gene transfer. All the cells used in this study were grown in DMEM medium supplemented with 10% fetal calf serum, at 37 °C, and under standard conditions. Synthetic siRNAs targeting human RD^{*INK4/ARF*} (siRNA-hRD; 5'-AGUCUUAACAGGAGGCAAUU-3', 5'-GAGAACCGCAA GUUAUGGAUU-3' and 5'-ACCCACUUUGUCAGGUUUCUU-3'), or siRNA-luciferase²⁴ as control, were transfected using Oligofectamine (Invitrogen) in accordance with the manufacturer's protocol. Briefly, 6 × 10⁶ HEK-293T cells (in a 10-cm-diameter dish, 75% confluency) were transfected with a mixture containing 0.8 nmol of each siRNA (higher amount in Fig. 1b) or 0.3 nmol (lower amount). Transfections were analysed 48 h after transfection. Retroviral constructs expressing shRNAs targeting either human RD^{*INK4/ARF*} (shRNA-hRD; see sequences above) or mouse RD^{*INK4/ARF*} (shRNA-mRD; 5'-GCACCA-GCACCAACCCGAGTGTATT-3' and 5'-GCTGTAGCAACAGTTGT AACAC-3') were cloned into pMSCV-puro (Clontech). Cdc6 was ectopically expressed from retroviral vector pLPC-puro, or tagged pcDNA-HA; Orc2 from tagged pCMV-V5; and oncogenic Ras (H-rasV12) from retroviral vector pLPC-puro. All the transfections into HEK-293T cells were performed according to standard procedures using Lipofectamine2000 (Invitrogen) and transfecting 20 µg of plasmid DNA (in those cases with two transfected amounts, these amounts correspond to 10 µg and 20 µg) into 6 × 10⁶ cells (in a 10-cm-diameter dish, ~75% confluent). Retroviral transductions were performed according to standard procedures. Retroviral supernatants were obtained from transfections of packaging HEK-293T cells performed with 20 µg of plasmid DNA (or with 10 µg and 20 µg when two amounts are used, as in Fig. 1b).

ChIP assays. Cells were crosslinked with a final concentration of 1% formaldehyde for 15 min at room temperature, and crosslinking was stopped by addition of glycine to a final concentration of 0.125 M. Crosslinked cells were lysed in buffer containing 1% SDS, 10 mM EDTA, 50 mM Tris-HCl pH 8.0. Lysates (400 µl at 1 µg protein per µl) were diluted 1:3 with 1% Triton-X100, 2 mM EDTA, 150 mM NaCl and 20 mM Tris-HCl (pH 8.0) containing protease inhibitors, and precleared with salmon sperm DNA/protein A agarose slurry (Upstate). The antibodies used for the immunoprecipitation were: rabbit polyclonal antibody against H3K9me3 (Upstate); mouse monoclonal antibody against Cdc6 (Ab-2; Cell Signalling); rabbit polyclonal antibodies against Mcm2 and Mcm3, produced by B. Stillman's laboratory^{25,26}; mouse monoclonal antibody against HA epitope (12CA5; Babco); mouse monoclonal antibody against V5 epitope (Invitrogen); and rabbit polyclonal antibodies against acetylated histone H3 or acetylated histone H4 (Upstate). DNA from precipitated complexes was amplified by PCR. The primers used were: for human RD^{*INK4/ARF*}, primers 4a and 5a (Supplementary Table 1); for human *INK4b* intron, primers 17a and 17b (Supplementary Table 1). We used previously

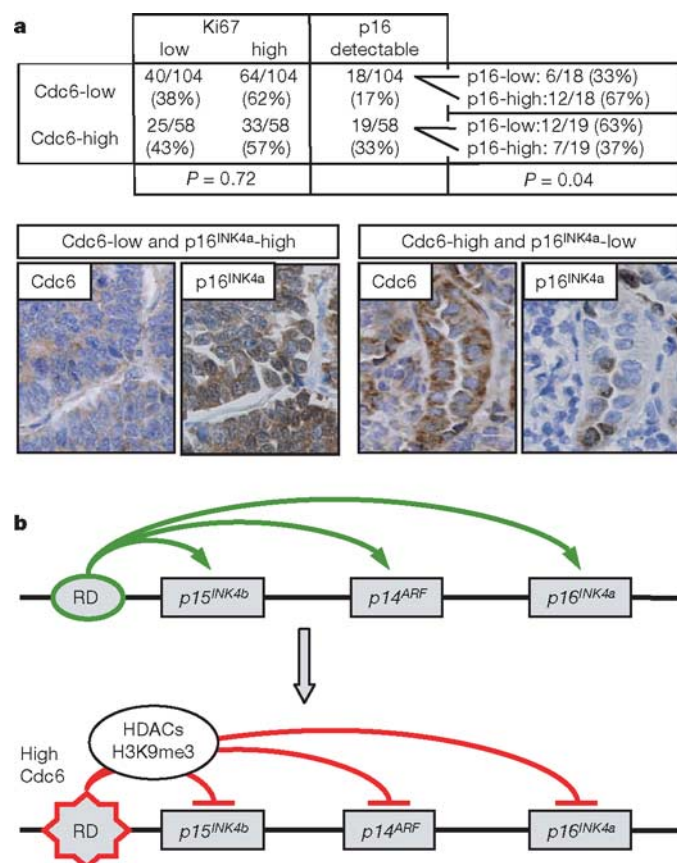


Figure 4 | Reciprocal relationship between Cdc6 and p16^{INK4a} protein levels in primary NSCLCs. **a**, Classification of a cohort of NSCLCs (*n* = 162) according to their levels of Cdc6 and p16^{INK4a} as measured by immunohistochemistry. Tumours with p16 detectable (>1% of positive nuclei) are subdivided into p16-low (1–25% of positive nuclei) or p16-high (>25% of positive nuclei). The stainings for two representative tumours are shown below. **b**, Mechanistic model of the oncogenic activity of Cdc6 through repression of the *INK4/ARF* locus.

reported primers for the following human sequences: *ARF* promoter²⁷, *p16^{INK4a}* promoter²⁷, *p73* gene²⁴ and lamin B2 replication origin¹⁴. For the following murine sequences we used: for RD^{INK4a/ARF}, 5'-TTCCTATTTCGCTGTAGCAAC-3' and 5'-AACTAACCA GGCCTCCTCCCA-3'; for *ARF* promoter, 5'-GCCTCG CCGATCTTCTA TTTTCT-3' and 5'-CCCATCGCGGTGACAGC-3; and for *p16^{INK4a}* promoter, 5'-CAGATTGCCCTCCGATGACTTC-3' and 5'-TGGA CCCGCACAGCAAAG AAGT-3'. Inputs correspond to PCR reactions using 1% of the total chromatin extracts used in the immunoprecipitation reactions.

Human samples. Samples of non-small cell lung carcinomas were obtained through the CNIO Tumour Bank Network.

All other assays were performed according to standard procedures and are detailed in Supplementary Information.

Received 31 October 2005; accepted 18 January 2006.

- Lowe, S. W. & Sherr, C. J. Tumor suppression by Ink4a-Arf: progress and puzzles. *Curr. Opin. Genet. Dev.* **13**, 77–83 (2003).
- Sherr, C. J. The INK4a/ARF network in tumour suppression. *Nature Rev. Mol. Cell Biol.* **2**, 731–737 (2001).
- Karakaidos, P. *et al.* Overexpression of the replication licensing regulators hCdt1 and hCdc6 characterizes a subset of non-small-cell lung carcinomas: synergistic effect with mutant p53 on tumour growth and chromosomal instability—evidence of E2F-1 transcriptional control over hCdt1. *Am. J. Pathol.* **165**, 1351–1365 (2004).
- Semple, J. W. & Duncker, B. P. ORC-associated replication factors as biomarkers for cancer. *Biotechnol. Adv.* **22**, 621–631 (2004).
- Murphy, N. *et al.* p16INK4A, CDC6, and MCM5: predictive biomarkers in cervical preinvasive neoplasia and cervical cancer. *J. Clin. Pathol.* **58**, 525–534 (2005).
- Fox, C. A. & McConnell, K. H. Toward biochemical understanding of a transcriptionally silenced chromosomal domain in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **280**, 8629–8632 (2005).
- Pennacchio, L. A. & Rubin, E. M. Genomic strategies to identify mammalian regulatory sequences. *Nature Rev. Genet.* **2**, 100–109 (2001).
- Cvetc, C. & Walter, J. C. Eukaryotic origins of DNA replication: could you please be more specific? *Semin. Cell Dev. Biol.* **16**, 343–353 (2005).
- Antequera, F. Genomic specification and epigenetic regulation of eukaryotic DNA replication origins. *EMBO J.* **23**, 4365–4370 (2004).
- Kawasaki, H. & Taira, K. Induction of DNA methylation and gene silencing by short interfering RNAs in human cells. *Nature* **431**, 211–217 (2004).
- Morris, K. V., Chan, S. W., Jacobsen, S. E. & Looney, D. J. Small interfering RNA-induced transcriptional gene silencing in human cells. *Science* **305**, 1289–1292 (2004).
- Serrano, M., Lin, A. W., McCurrach, M. E., Beach, D. & Lowe, S. W. Oncogenic ras provokes premature cell senescence associated with accumulation of p53 and p16INK4a. *Cell* **88**, 593–602 (1997).
- Stucki, M., Stagljar, I., Jonsson, Z. O. & Hubscher, U. A coordinated interplay: proteins with multiple functions in DNA replication, DNA repair, cell cycle/checkpoint control, and transcription. *Prog. Nucleic Acid Res. Mol. Biol.* **65**, 261–298 (2001).
- Abdurashidova, G. *et al.* Localization of proteins bound to a replication origin of human DNA along the cell cycle. *EMBO J.* **22**, 4294–4303 (2003).
- Gonzalez, M. A., Tachibana, K. E., Laskey, R. A. & Coleman, N. Control of DNA replication and its potential clinical exploitation. *Nature Rev. Cancer* **5**, 135–141 (2005).
- Frolova, N. S., Schek, N., Tikhmyanova, N. & Coleman, T. R. *Xenopus* Cdc6 performs separate functions in initiating DNA replication. *Mol. Biol. Cell* **13**, 1298–1312 (2002).
- Tao, L., Dong, Z., Leffak, M., Zannis-Hadjopoulos, M. & Price, G. Major DNA replication initiation sites in the c-myc locus in human cells. *J. Cell. Biochem.* **78**, 442–457 (2000).
- Araujo, F. D. *et al.* Identification of initiation sites for DNA replication in the human *dnmt1* (DNA-methyltransferase) locus. *J. Biol. Chem.* **274**, 9335–9341 (1999).
- Ladenburger, E. M., Keller, C. & Knippers, R. Identification of a binding region for human origin recognition complex proteins 1 and 2 that coincides with an origin of DNA replication. *Mol. Cell. Biol.* **22**, 1036–1048 (2002).
- Mutskov, V. & Felsenfeld, G. Silencing of transgene transcription precedes methylation of promoter DNA and histone H3 lysine 9. *EMBO J.* **23**, 138–149 (2004).
- Katan-Khaykovich, Y. & Struhl, K. Heterochromatin formation involves changes in histone modifications over multiple cell generations. *EMBO J.* **24**, 2138–2149 (2005).
- Wistuba, I. I., Gazdar, A. F. & Minna, J. D. Molecular genetics of small cell lung carcinoma. *Semin. Oncol.* **28**, 3–13 (2001).
- Delgado, S., Gomez, M., Bird, A. & Antequera, F. Initiation of DNA replication at CpG islands in mammalian chromosomes. *EMBO J.* **17**, 2426–2435 (1998).
- Gonzalez, S., Prives, C. & Cordon-Cardo, C. p73 α regulation by Chk1 in response to DNA damage. *Mol. Cell. Biol.* **23**, 8161–8171 (2003).
- Mendez, J. & Stillman, B. Chromatin association of human origin recognition complex, cdc6, and minichromosome maintenance proteins during the cell cycle: assembly of prereplication complexes in late mitosis. *Mol. Cell. Biol.* **20**, 8602–8612 (2000).
- Ekholm-Reed, S. *et al.* Dereglulation of cyclin E in human cells interferes with prereplication complex assembly. *J. Cell Biol.* **165**, 789–800 (2004).
- Arcellana-Panlilio, M. Y. *et al.* Decreased expression of the INK4 family of cyclin-dependent kinase inhibitors in Wilms tumor. *Genes Chromosom. Cancer* **29**, 63–69 (2000).
- Wei, W., Hemmer, R. M. & Sedivy, J. M. Role of p14^{ARF} in replicative and induced senescence of human fibroblasts. *Mol. Cell. Biol.* **21**, 6748–6757 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements S.G. was supported by the Human Frontiers Science Program Organization and by the FIS from the Spanish Ministry of Health. Research was supported by the CNIO and by grants from the Spanish Ministry of Education and Science (to M.S., F.A. and J.M.), the European Union project INTACT (to M.S.) and Fundacion Caja Madrid (to J.M.).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.S. (mserrano@cnio.es).

RETRACTION

doi:10.1038/nature04622

Magnetic carbon

T. L. Makarova, B. Sundqvist, R. Höhne, P. Esquinazi, Y. Kopelevich, P. Scharff, V. Davydov, L. S. Kashevarova & A. V. Rakhmanina

Nature **413**, 716–718 (2001)

In this Letter we reported high-temperature ferromagnetism in a polymeric phase of pure carbon that was purportedly free of ferromagnetic impurities¹. Since then, however, measurements made on the same and similar samples using particle-induced X-ray emission (PIXE) with a proton microbeam have indicated that these had considerable iron content^{2–4}. Also, polymerized C₆₀ samples mixed with iron before polymerization had a similar Curie temperature (500 K) to those we described¹, owing to the presence of the compound Fe₃C (cementite)⁵. In addition, it has since been shown that the pure rhombohedral C₆₀ phase is not ferromagnetic⁶.

Nevertheless, magnetic order in impurity-free graphitic structures at room temperature has been demonstrated independently⁷ (before and after publication of ref. 1). Ferromagnetic properties may yet be found in polymerized states of C₆₀ with different structural defects and light-element (H, O, B, N) content⁸.

T.L.M. and P.S. decline to sign this retraction because they do not believe that the earlier results¹, supported in subsequent studies^{9,10}, are totally invalidated by these findings^{2–6}.

1. Makarova, T. L. *et al.* Magnetic carbon. *Nature* **413**, 716–718 (2001); corrigendum **436**, 1200 (2005).
2. Höhne, R. & Esquinazi, P. Can carbon be ferromagnetic? *Adv. Mater.* **14**, 753–756 (2002).
3. Spemann, D. *et al.* Evidence for intrinsic weak ferromagnetism in a C₆₀ polymer by PIXE and MFM. *Nucl. Instrum. Meth. B* **210**, 531–536 (2003).
4. Han, K. H. *et al.* Observation of intrinsic magnetic domains in C₆₀ polymer. *Carbon* **41**, 785–795 (2003); addendum **431**, 2425–2426 (2003).
5. Talyzin, A., Dzwilewski, A., Drubovinsky, L., Setzer, A. & Esquinazi, P. Magnetic properties of polymerized C₆₀ with Fe. Preprint at (<http://arXiv.org/abs/cond-mat/0602306>).

6. Boukhalov, D. W. *et al.* Testing the magnetism of polymerized fullerene. *Phys. Rev. B* **69**, 115425 (2004).
7. Makarova, T. & Palacio, F. (eds) Ch. 24 *Carbon-based Magnetism* (Elsevier, Amsterdam, 2006).
8. Makarova, T. L. *et al.* Magnetism in photopolymerized fullerenes. *Carbon* **41**, 1575–1584 (2003).
9. Narozhnyi, V. N. *et al.* Ferromagnetic carbon with enhanced Curie temperature. *Physica B* **329**, 1217–1218 (2003).
10. Wood, R. A. *et al.* Ferromagnetic fullerene. *J. Phys. Condens. Matter* **14**, L385–L391 (2002).

RETRACTION

doi:10.1038/nature04621

Evidence for spin-charge separation in quasi-one-dimensional organic conductors

T. Lorenz, M. Hofmann, M. Grüninger, A. Freimuth, G. S. Uhrig, M. Dumm & M. Dressel

Nature **418**, 614–617 (2002)

In this Letter, we described the anomalous behaviour of the thermal conductivity k in quasi-one-dimensional Bechgaard salts. As we have since observed a conventional temperature dependence of k in another quasi-one-dimensional organic conductor, TTF-TCNQ, we remeasured the value of k in our Bechgaard salts and identified a heat leakage in the original experimental set-up, which becomes relevant for samples with very small values of k at low temperatures. In an improved set-up, the low-temperature maxima of k (as reported in Figs 2 and 3) are not reproduced, indicating that our estimate of the phonon contribution to the thermal conductivity of the Bechgaard salts is no longer valid. Because a reliable decomposition of k into a magnetic and a phononic contribution was the basis of our interpretation, we have to withdraw the conclusion that our data provided evidence for spin-charge separation.

CORRIGENDUM

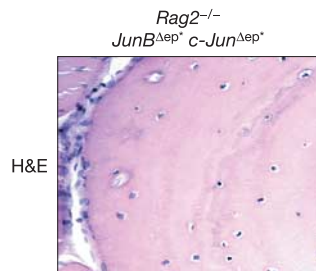
doi:10.1038/nature04620

Psoriasis-like skin disease and arthritis caused by inducible epidermal deletion of Jun proteins

Rainer Zenz, Robert Eferl, Lukas Kenner, Lore Florin,
Lars Hummerich, Denis Mehic, Harald Scheuch, Peter Angel,
Erwin Tschachler & Erwin F. Wagner

Nature 437, 369–375 (2005)

Mislabelling of a histological section caused the wrong micrograph of a wild-type control to be inserted in Fig. 5b (bottom panel, middle column) of this Article: as published, this is identical to the control bone picture shown in Fig. 3c (top left panel). The panel that should have been shown in Fig. 5b (see below) is from the *Rag2*-mutant mouse; the figure legend remains the same. The mix-up with the control figure does not affect the validity of the data or our original conclusions.



-  FOCUS
-  SPOTLIGHT
-  RECRUITMENT
-  ANNOUNCEMENTS
-  EVENTS

naturejobs

Cracking the tax code

Apart from being the day designated for practical jokes, 1 April offers another reason for heightened paranoia — it means that there are just two weeks left for filing US tax forms. But some US postdocs may have reason to celebrate, rather than fear, this deadline.

According to a survey of postdocs by the scientific research society Sigma Xi, of Research Triangle Park, North Carolina, about half the recipients of National Research Service Awards (NRSAs) who were queried turned out to be paying thousands of dollars in unnecessary social-security contributions. As a result they may qualify for refunds.

How has this happened? It is all down to the changing tax status of fellowship funding. About 25 years ago, the US Internal Revenue Service said that stipends counted as taxable income. That decision was reversed in the early 1980s, but not all universities — which administer this federal money — got the message. As a result, the tax status of postdocs has been assessed in one of three different ways.

If the stipend is seen as a wage, then the institution generally withholds social-security contributions. If the

university views the fellows as independent contractors, then it doesn't withhold social-security contributions but subjects the fellows to self-employment tax. And if the university treats fellows as fellows, the postdocs have no social-security tax liability.

Nearly half of the NRSA fellows surveyed by Sigma Xi were treated — erroneously — under the first category, and another 15% were wrongly taxed as independent contractors. If these figures are extrapolated to all 1,500 NRSA fellows, then collectively they have been overtaxed by about \$2.8 million per year, says Geoff Davis, author of Sigma Xi's survey.

So now is a good time for NRSA fellows to check their pay stubs to see whether they have had social-security taxes withheld. If the joke's been on them, then it's time to turn the tables and have the last laugh.



Paul Smaglik, Naturejobs editor

CONTACTS

Publisher: Ben Crowe
Editor: Paul Smaglik
Assistant Editor: Corie Lok

European Head Office, London
 The Macmillan Building,
 4 Crinan Street
 London N1 9XW, UK
 Tel: +44 (0) 20 7843 4961
 Fax: +44 (0) 20 7843 4996
 e-mail: naturejobs@nature.com

European Sales Manager:
 Andy Douglas (4975)
Natureevents: Silke Opstrup (4994)

UK/RoW/Ireland/Italy:
 Nils Moeller (4953)
 Irene Viglia-Atton (4944)
Scandinavia/Spain/Portugal:
 Evelina Rubio Håkansson (4973)
France/Switzerland/Belgium:
 Amelie Pequignot (4974)
**Germany/Austria/
 The Netherlands:**
 Reya Silao (4970)

Advertising Production Manager:
 Billie Franklin
 To send materials use London
 address above.
 Tel: +44 (0) 20 7843 4814

Fax: +44 (0) 20 7843 4996
 e-mail: naturejobs@nature.com
Naturejobs web development: Tom Hancock
Naturejobs online production: Niamh Shields

**European Satellite Office
 Germany/Austria/
 The Netherlands:**
 Patrick Phelan
 Tel: +49 89 54 90 57 11
 Fax: +49 89 54 90 57 20
 e-mail: p.phelan@nature.com

US Head Office, New York
 75 Varick Street,
 9th Floor, New York,

NY 10013-1917
 Tel: +1 800 989 7718
 Fax: +1 800 989 7103
 e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

Japan Head Office, Tokyo
 Chiyoda Building,
 2-37 Ichigayatamachi,
 Shinjuku-ku,
 Tokyo 162-0843
 Tel: +81 3 3267 8751
 Fax: +81 3 3267 8746
Asia-Pacific Sales Director: Rinoko Asami
 e-mail: rasami@naturejpn.com

Written in the blood

Research into angiogenesis has survived the 1990s hype about an imminent cure for cancer, and shows promising results in many areas — but don't tell the newspapers, says **Ricki Lewis.**

Judah's going to cure cancer in two years." Those words, spoken casually by Nobel laureate James Watson to a reporter at a banquet, were trumpeted on the front page of *The New York Times* in 1998. Judah Folkman, a researcher at Harvard University and Children's Hospital Boston, suddenly found himself in the spotlight. There was a storm of interest in his field, angiogenesis, which studies how blood vessels are formed and how this process might be harnessed to tackle a host of diseases.

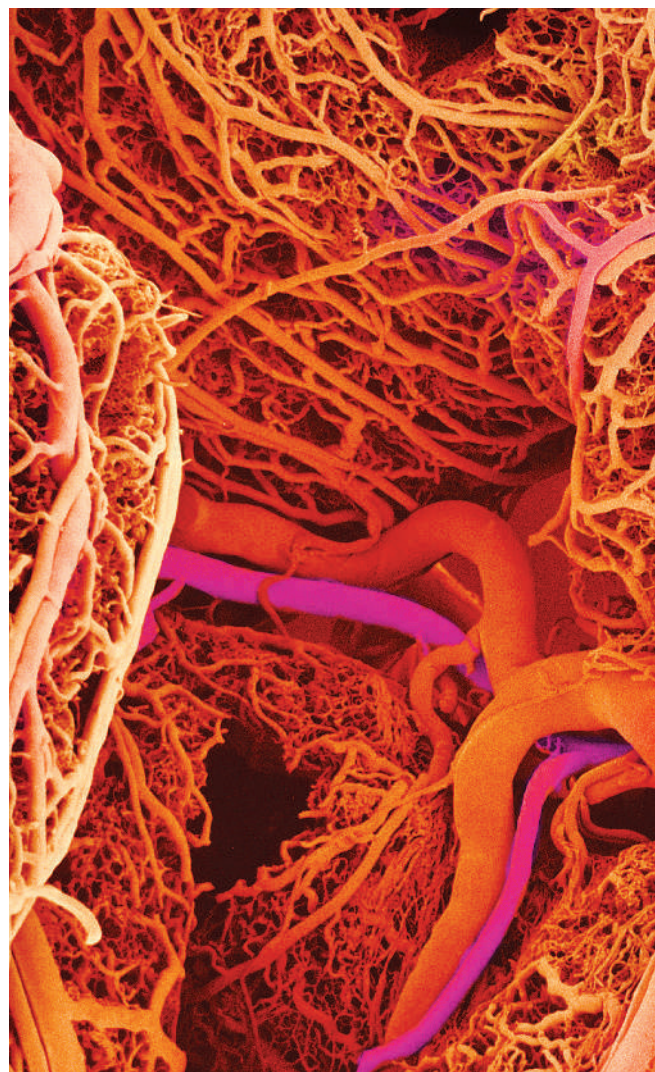
Eight years later, the hype has waned and cancer is still uncured. But interest remains in a field that still offers therapeutic promises. Folkman had first proposed stifling angiogenesis to starve tumours of their blood supply in 1971, and after *The New York Times* headline, the quest for inhibitors kicked into high gear.

Molecules that could regulate blood-vessel formation have implications way beyond cancer, too. Manipulating the ebb and flow of vessels that feed our tissues could equally well improve the supply of healing oxygen to tissues starved by heart disease or damaged by injury or skin disease. Angiogenesis is implicated in at least 70 disorders, with a worldwide potential market of 500 million patients. "Blood vessels are everywhere, and so many diseases are affected by angiogenesis," says Peter Carmeliet, adjunct director of the Centre for Transgene Technology and Gene Therapy at the Catholic University of Leuven, Belgium.

Altering angiogenesis is a huge field that has already proved its worth in many areas, says William Li, president of the Angiogenesis Foundation in Cambridge, Massachusetts. More than 2,500 laboratories and 375 companies worldwide are engaged in angiogenesis research and development (R&D), Li notes. "All signs point to continued astronomical growth," he says, "because blood vessels play a sentinel role in both physiology and pathology. It is a ripe target for fundamental research as well as drug discovery."

The New York Times story triggered a host of clinical trials: 56 for anti-angiogenics, to tackle cancer, and 21 for pro-angiogenics, to combat heart disease. "The expectations were enormous," Carmeliet remembers. Great hopes were aroused by the development of a humanized monoclonal antibody (the future Genentech drug Avastin, or bevacizumab) to block the production of vascular endothelial growth factor (VEGF), which is key to creating the vessels that supply blood to tumours. Disappointingly, it turned out not to extend survival in colon cancer without additional chemotherapy. But this apparent setback broadened research beyond what Gregory Plowman, senior director of tumour biology at Genentech in South San Francisco, calls 'the VEGF axis'. "It is the most validated target, so we are looking at every modification of that system," he says, as well as other growth factors and pathways.

Today, angiogenesis R&D continues on the two tracks of inhibition and stimulation, says Harold



Strong connection: controlling the growth of blood vessels could help treat a wide range of diseases.



"Blood vessels are everywhere, and so many diseases are affected by angiogenesis"

— Peter Carmeliet

Dvorak, a pathologist at Harvard Medical School. Of the 75 current government-sponsored clinical trials in the United States for angiogenesis-related drugs, 57 target cancer and 10 cardiovascular disease. The oncology emphasis reflects the fact that it is easier to demolish the disorganized blood vessels of a tumour than to recreate the complex network damaged by heart disease. "Extensive programme controls, involving many growth factors and their inhibitors, have to act in a particular sequence to make normal blood vessels," explains Dvorak.

Teams and networks

The job market in angiogenesis is a vast continuum, from the small biopharmaceutical firms that often pioneer drugs and technologies, to big drug firms, with so many collaborations that researchers can "swing from vine to vine in the industry", Li says. And the first wave of successful angiogenesis drugs has encouraged the development of technologies to track the drug response, Li adds.

The field requires skills both integrative and eclectic. In addition to basic competence in cell- and molecular-biology techniques, a job may entail a specific application, such as dissecting eyeballs or



interpreting brain scans. "There is no straight path to follow in the angiogenesis field," says Plowman. For example, the company hired several neurobiologists, because some of the aspects of sprouting axons resemble the sprouting of blood vessels in angiogenesis.

The story of Macugen (pegaptanib sodium), a drug developed by OSI Pharmaceuticals of Melville, New York, to treat the eye disease age-related macular degeneration (AMD), illustrates the diversity of skills behind developing an angiogenesis-targeting drug, as well as a typical corporate pedigree. "Macugen was isolated from a large combinatorial library of random sequences," recalls Barry Polisky, now chief scientific officer at Sirna Therapeutics in Boulder, Colorado, who co-discovered the compound with Nebojsa Janjic, now chief scientific officer at Replidyne in Louisville, Colorado, while both were at nearby NeXstar Pharmaceuticals a decade ago. Janjic and Polisky pursued AMD, rather than cancer, because it is a more specific target, with simpler delivery and a huge market. After discovering that attaching polyethylene glycol improved the stability of the compound in the eye, the researchers needed help. "We were not terribly knowledgeable about angiogenesis, so we looked for people with animal models. We found a group at Harvard, and they tested the compound that became Macugen," says Polisky.



Michael Detmar: seeks a number of different skills.



Joseph Frank: wants a clear picture of vessel formation.

Networks of angiogenesis labs streamline job searching. In Europe, two such groups parallel the cancer and cardiovascular markets. The 16 research centres of the European Union-sponsored Stroma Consortium represent seven nations. Rather than inhibiting angiogenesis, their cancer strategy pursues antibodies against tumour markers, says coordinator Raffaella Giavazzi, of the Mario Negri Institute for Pharmacological Research in Bergamo, Italy. The European Vascular Genomics Network represents 25 basic and clinical research facilities exploring various angiogenesis-based approaches to treating cardiovascular disease. The network offers postdoctoral positions, with frequent cross-fertilization among participating laboratories.

A more localized collaboration with an interdisciplinary approach to angiogenesis in brain tumours includes investigators from Harvard, the Massachusetts Institute of Technology and Massachusetts General Hospital. "We investigate cancer as a complex, dynamic biosystem using an interdisciplinary approach that combines computation with experiment," explains Thomas Deisboeck, director of the complex-biosystems modelling laboratory. The programme seeks PhD scientists who have a strong theoretical background with "a zest for interdisciplinary research, for being part of a highly collaborative dynamic environment, and for thinking outside the box," he adds.

Space for specialists

A researcher can also specialize within the field of angiogenesis: in technique, aspect of the process or even body part. For example, an intramural National Institutes of Health (NIH) postdoctoral programme focuses on magnetic resonance imaging and microscopy. "The research involves direct visualization of cells that participate in new vessel formation," says Joseph Frank, chief of the experimental neuroimaging section in the NIH's laboratory of diagnostic radiology research. And an extramural NIH pre- and postdoctoral programme at Albany Medical Center in New York investigates the role of extracellular matrix remodelling in angiogenesis in cardiovascular disease.

Michael Detmar, professor of pharmacogenomics at the Swiss Federal Institute of Technology in Zurich, focuses on skin. His lab uses mouse models to investigate cancers and wound healing. "We are interested in scientists with experience in genetics, bioinformatics and developmental biology," Detmar says.

Research positions in angiogenesis are also available at bachelor's or master's degree levels. At Regeneron Pharmaceuticals in Tarrytown, New York, research associates need a BSc or MSc in biology, chemistry, pharmacology or biochemistry, with cell- and molecular-biology skills. At Genentech, research associates should have experience in microarrays, data gathering, cell culture and human tumour xenografts. PhD-level positions in industry require experience in clinical-trial design and drug submission.

The future of angiogenesis research and drug development is rosy, with the VEGF axis ripe for exploiting in many disorders. Although there are several routes to a career in angiogenesis, Genentech's Plowman lists two additional skills. "Enthusiasm and passion," he says. "You can't train for that."

Ricki Lewis is a freelance writer in Scotia, New York.

MOVERS

Scott Hubbard, Carl Sagan chair for the study of life in the Universe, SETI Institute, Mountain View, California



2002-06: Director, NASA Ames Research Center, Moffett Field, California

2001-02: Deputy director for Research, NASA Ames Research Center

2000-01: Mars programme director, NASA, Washington DC

1999-2000: Associate director for astrobiology and space programmes, NASA Ames Research Center

Scott Hubbard's interests in space exploration were triggered by Sputnik's launch in 1957, a predilection for Isaac Asimov's science fiction, and some backyard rocket experiments that went awry. Seeing this stream of man-made objects flying through the sky set off Hubbard's imagination and his career.

With a degree in physics-astronomy from Vanderbilt University in Tennessee, he found a job at the Lawrence Berkeley National Laboratory (LBNL) that let him pursue graduate studies at the University of California, Berkeley.

While at LBNL, Hubbard invented a novel radiation-detection technology that ultimately found application in space missions. It was also the basis for a small company Hubbard founded with his colleagues in 1980. Becoming a general manager of the start-up taught him not only management and leadership skills, which he would hone throughout his career, but also the art of doing science. "Making the judgement between when to be exquisitely careful and knowing when quick and dirty is good enough is the hallmark of good research," he says.

After selling the business, he eventually landed at NASA, where he became manager of the successful Lunar Prospector mission. It was Hubbard who answered the call for a simpler, cheaper way to get to the surface of Mars. His solution was to use airbags to cushion the landing of a probe, now a standard procedure in space-exploration missions. Hubbard also helped create NASA's new interdisciplinary Astrobiology Institute devoted to studying life in the Universe. "I always wanted to explore space — at least with my imagination," he says.

Although he cites imagination as his most valuable skill, his leadership abilities have consistently been in demand. Hubbard faced his two most challenging roles while at NASA: rebuilding the Mars programme after two failed missions, and then serving as the NASA representative on the Columbia Accident Investigation Board after the shuttle disaster in 2003. The Mars programme he put together for the Odyssey, Spirit and Opportunity missions is still in place today, and he directed the experimental analyses that determined why Columbia was lost.

Hubbard returns to research in his dream job at the Search for Extraterrestrial Intelligence (SETI) Institute. His goals are to strengthen SETI's research capabilities and to carry on Carl Sagan's legacy of communicating to the public why space exploration is so exciting. ■

Virginia Gewin

MENTORS & PROTÉGÉS

Emerging into the light

A troubling feature of South Africa's research landscape is the ageing profile of the top researchers, which highlights the need for the transfer of skills to a new generation. Meanwhile, inexperienced young scientists are struggling in an increasingly competitive environment.

To address this problem, the University of Cape Town implemented in 2003 an institution-wide mentoring programme, called the Emerging Researcher Programme (ERP). Junior faculty members are paired up, for a defined time, with retired colleagues chosen for their outstanding research record and proven mentoring skills.

Open to all permanent staff, the ERP began with 46 self-identified 'emerging researchers'. Within three years, that number has grown to 170, about one-third of whom are scientists.

The idea is that, with appropriate mentoring, inexperienced staff will develop the skills and self-confidence to become mature researchers, capable of doing sustained, high-quality work.

The participants get an appropriate mentor for assistance with specific tasks — preparing an article for publication, for example. Meetings take place approximately twice a month until the task is complete.

In addition to the individual sessions, our mentors offer a series of seminars

covering topics such as writing grant proposals, project management, writing for publication, and career planning.

The university also provides an annual budget of 1 million rand (US\$157,000) for enlisted emerging researchers, guaranteeing modest funding for up to three years. The award process is competitive, requiring well-crafted proposals and committee evaluation, with post-award monitoring.

Unlike many mentoring situations, the ERP is structured and needs-driven. Formalizing the programme ensures that all participants have access to the same quality of mentoring, and makes it easy to monitor progress. Details of the participation level of both mentor and protégé, along with ensuing research output, are recorded in a database from which regular reports are generated.

The success of the ERP is measured against a set of key performance indicators, including peer-review publication, successful external funding proposals and promotions of the participating young researchers. Early indications are that the programme is achieving its goal. For example, in 2005 there were 17 promotions and a significant increase in publications. ■

Lyn Holness is the manager of the Emerging Researcher Programme at the University of Cape Town, South Africa.

GRADUATE JOURNAL

Off the straight and narrow

On the final leg of a PhD, your career options can seem claustrophobically narrow, with only the research track ahead. So how do you explore your options?

My solution was to enrol in a mentoring programme, organized by my labour union. Students are assigned a senior colleague, who acts as a mentor as you map the employment terrain, analyse skills and interests, and build networks. Along with the gratification of helping a younger colleague, mentors may gain from catching up on what's happening in science, or reflecting on their own career choices in a new way.

Through my mentor, I've met biologists working in fields from publishing and industry to legislation and government. At first, it felt a little embarrassing to admit to these successful professionals that I have no clue about my future, but the response has been sympathetic. Indeed, many seemed surprised by where they themselves had ended up. One woman was offered her job because the employer remembered her from a previous application — for a different post. Similarly, a high-ranking administrative career began as summer field work.

Paradoxically, the stories have taught me to plan less, rather than more. For a graduate student with little work experience outside university, the best idea is not to restrict yourself to one path, but be open enough to grasp unexpected opportunities. With a little exploration, the possibilities are endless. ■

Katja Bargum is a graduate student in the Department of Biological and Environmental Sciences, University of Helsinki, Finland.

Party smart card

A transfer of allegiance.

Barrington J. Bayley

Pig farmer Igor scratched his porky belly under his smock on emerging from the sty after tipping swill into the trough. Behind him were satisfied grunts from feeding snouts.

As a member of the Smallholders' Alliance Party, he felt supremely confident to be treading the muck of the farmyard. The Party would make life better for all pig farmers...but what was this? Sacha, his brat of a nephew foisted on him after the murder of his family, had fished his damned books out of the slurry bin where Igor had thrown them, and was wiping the bemerded covers. Books! No use to a pig farmer. Igor hurried to cuff him. The ten-year-old went sprawling in the filth, the volumes flopping from his hands. Scarcely able to read, Igor did not take in the titles: *Weapons of the Great Patriotic War* and *Nuclear Power: What it Means*. Sacha stared up at him with hard, steady eyes. Igor hated the boy. No matter how much you thrashed him you couldn't break his spirit.

"Get to the kitchen. Mash more swill."

Sacha stood up, not retrieving his books. He knew it would only enrage Igor the more. Unlike Igor, he lacked a Party transfer on his brow. Such wasn't for the likes of him.

He trudged off. Something else distracted Igor. Three horsemen were descending the eastern slope. He stood nervously as they trotted into the farmyard. They were shabbily dressed, their mounts ungroomed, but the coloured transfers on their brows made Igor afraid. They were not crude and smudged like his, but well defined: a blue circle and a white bar against a red background.

The leader spoke down at him. "I am hetman of the Sovrina Party. We are here to enrol you."

"I am loyal to the Smallholders' Alliance Party!" Igor shouted. He tried to run, but another Sovrina Party man had dismounted and now seized him. "Huh!" the hetman snorted. "Just look at that transfer. Blurred and wishy-washy. Brewed in a cowshed."

There was something in his hand. It was a party decal card. He leaned down, slapping it on Igor's brow. With a cold sensation the card's thin transfer film attached itself. It was stronger than the Smallholders' Alliance. Already the weaker transfer was obliterated. New psychoactive opin-



ion-bearing vector molecules migrated through Igor's skin, his skull, his meninges, and flooded his brain. In a few exhilarating moments all his political views were rearranged. Sovrina! A country that would rule for hundreds of kilometres around! His country!

Then a bugle-call made the hetman look round. "Damn!" he muttered. "They've found us."

A second troop was cantering down the slope, this one ten strong. With whoops of triumph they circled round those in the farmyard then reined in, horses stamping. The leader was easy to pick out: moustachioed, wearing a sheepskin jacket and an Astrakhan hat and brandishing a Kalashnikov, which made Igor's eyes bulge. Their brow transfers were spectacular, displaying a collage of a tank, a warplane and a missile.

"Get down," the leader ordered his men.

Ivan Ivorschenko, self-styled Servant of the Great, lounged in his saddle, watching his men slap Forced Reconstruction Party decals on the prisoners. It was ironic that the collapse of the world's population, due to disease and famine, while emptying the cities and bringing large-scale organization to a calamitous end, had allowed small technology to come into its own. Psychoactive decals needed no factories. They could be made in a kitchen, if you knew how. The world was a morass, a welter of tiny parties. Transfer after transfer must have been forced on hapless peasants like the pig farmer. Those who made the strongest decals would prevail eventually, as humanity slowly pulled itself together.

His forehead itched. He wiped it with his sleeve. His temporary transfer smeared off (only he among his troop had a temporary one). It revealed a brilliant, flashing mushroom cloud, the strongest psychotransfer of all, one which could not be expunged and which overrode all others. The party card of the Warmaster, prophesied to appear one day and take charge of the Forced Reconstruction Party, according to the elite cabal Ivorschenko belonged to, restoring everything. Tanks! Submarines! Warplanes! Atomic energy!

Even ammunition for his ancient Kalashnikov.

Ivorschenko flinched as a crossbow bolt was put through the head of the pig farmer. For some reason his transfer wasn't taking effect properly. Only then did Ivorschenko notice that a small boy had crept from the miserable cottage and was picking up books from the evil-smelling mud.

"What have you got there?"

The boy approached, allowing Ivorschenko to read the titles. Ivorschenko smiled.

"Well done, boy. You've got the right idea. Here, take this."

He reached into his saddlebag and handed down a Party booklet blotchily printed on coarse paper. The boy's eyes gleamed on seeing the title: *How to Rule the World with Ballistic Missiles*.

"What's your name?"

"Sacha, sir."

"And him?" Ivorschenko nodded to the smock-clad corpse lying in the muck.

"He was no better than the pigs he kept."

"Then the Party was right to reject him."

There was something about the lad. His gaze was level and unafraid. Ivorschenko reached down and pulled him up on his horse. "Let me finish your education."

He didn't know why he did it. Warmaster decals were rare, and he only possessed one. He took it from his pocket and applied it to Sacha's brow. The boy's eyes shone with sudden transformation.

"Ride on!" he commanded in a piping voice.

Ivorschenko spurred his horse. He had found him!

The Warmaster!

Barrington J. Bayley's science-fiction production spans half a century. *The Soul of the Robot*, *The Zen Gun* (novels) and *Gnostic Endings* (a story collection) are typical of his work.